Federal Reserve Bank of St. Louis

REVIEW

NOVEMBER/DECEMBER 2009 VOLUME 91, NUMBER 6

Review is published six times per year by the Research Division of the Federal Reserve Bank of St. Louis and may be accessed through our website: **research.stlouisfed.org/publications/review**. All nonproprietary and nonconfidential data and programs for the articles written by Federal Reserve Bank of St. Louis staff and published in *Review* also are available to our readers on this website. These data and programs are also available through Inter-university Consortium for Political and Social Research (ICPSR) via their FTP site: www.icpsr.umich.edu/pra/index.html. Or contact the ICPSR at P.O. Box 1248, Ann Arbor, MI 48106-1248; 734-647-5000; **netmail@icpsr.umich.edu**.

Single-copy subscriptions are available free of charge. Send requests to: Federal Reserve Bank of St. Louis, Public Affairs Department, P.O. Box 442, St. Louis, MO 63166-0442, or call (314) 444-8809.

General data can be obtained through FRED (Federal Reserve Economic Data), a database providing U.S. economic and financial data and regional data for the Eighth Federal Reserve District. You may access FRED through our website: **research.stlouisfed.org/fred**.

ISSN 0014-9187

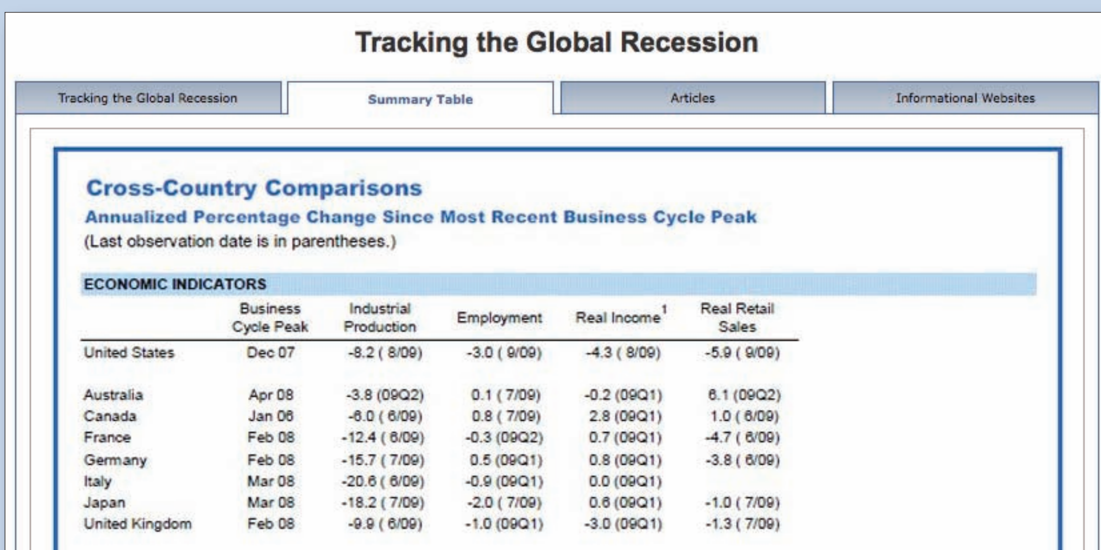
More Clarity and Data from the St. Louis Fed:

Walk step by step through the financial crisis with a glossary, links, and related research:

timeline.stlouisfed.org

Updated international data and commentary on the current downturn:

research.stlouisfed.org/recession/index.html



Tracking the Global Recession

Tracking the Global Recession | Summary Table | Articles | Informational Websites

Cross-Country Comparisons

Annualized Percentage Change Since Most Recent Business Cycle Peak

(Last observation date is in parentheses.)

ECONOMIC INDICATORS

	Business Cycle Peak	Industrial Production	Employment	Real Income[1]	Real Retail Sales
United States	Dec 07	-8.2 (8/09)	-3.0 (9/09)	-4.3 (8/09)	-5.9 (9/09)
Australia	Apr 08	-3.8 (09Q2)	0.1 (7/09)	-0.2 (09Q1)	6.1 (09Q2)
Canada	Jan 08	-8.0 (6/09)	0.8 (7/09)	2.8 (09Q1)	1.0 (6/09)
France	Feb 08	-12.4 (6/09)	-0.3 (09Q2)	0.7 (09Q1)	-4.7 (6/09)
Germany	Feb 08	-15.7 (7/09)	0.5 (09Q1)	0.8 (09Q1)	-3.8 (6/09)
Italy	Mar 08	-20.6 (6/09)	-0.9 (09Q1)	0.0 (09Q1)	
Japan	Mar 08	-18.2 (7/09)	-2.0 (7/09)	0.6 (09Q1)	-1.0 (7/09)
United Kingdom	Feb 08	-9.9 (6/09)	-1.0 (09Q1)	-3.0 (09Q1)	-1.3 (7/09)

Auctions as a Vehicle to Reduce Airport Delays and Achieve Value Capture

Jeffrey P. Cohen, Cletus C. Coughlin, and Lesli S. Ott

Congestion at airports imposes large costs on airlines and their passengers. A key reason for congestion is that an airline schedules its flights without regard to the costs imposed on other airlines and their passengers. As a result, during some time intervals, airlines schedule more flights to and from an airport than that airport can accommodate and flights are delayed. This paper explores how a specific market-based proposal by the Federal Aviation Administration (FAA), which includes the use of auctions to determine the right to arrive or depart in a specific time interval at airports in the New York City area, might be used as part of a strategy to mitigate delays and congestion. By explaining the underlying economic theory and key arguments with minimal technical jargon, the paper allows those with little formal training in economics to understand the fundamental issues associated with the FAA's controversial proposal. Moreover, the basics of the proposed auction process, known as a combinatorial auction, and value capture are also explained.
(JEL R41, R48, L93)

Federal Reserve Bank of St. Louis *Review*, November/December 2009, *91*(6), pp. 569-87.

In rankings of on-time arrivals and departures for major airports, New York City area airports fare poorly. For 32 major airports during 2008, the on-time arrivals of Newark Liberty International Airport (Newark) ranked 32nd, LaGuardia Airport (LaGuardia) ranked 31st, and John F. Kennedy International Airport (Kennedy) ranked 29th.[1] Their on-time arrival percentages—Newark's 62 percent, LaGuardia's 63 percent, and Kennedy's 69 percent—revealed that fewer than seven of every ten flights arrived on time.[2] Thus, more than three of every ten flights arrived 15 or more minutes past their scheduled arrival time.

Many factors, such as weather and mechanical problems, determine an airport's on-time performance. This paper focuses on the delays and congestion associated with the scheduling of flight arrivals and departures. Generally speaking, an airline schedules its flights without regard to the costs it imposes on other airlines and their passengers. When airlines schedule more flights to and from an airport than that airport can accommodate, flights are delayed. These delays impose large costs on the airlines themselves and, ultimately, on passengers. With respect to New York City area airports, Levine (2009) cites an estimate that these airports are involved in three-quarters of congestion delay in the entire U.S. system. We explore

[1] Using on-time departure data provides a similar picture. For 2008 Kennedy ranked 26th, LaGuardia ranked 28th, and Newark ranked 31st. See "Airline On-Time Statistics" (www.bts.gov/xml/ontimesummarystatistics/src/index.xml).

[2] These on-time arrival statistics likely understate the extent of the problem. Because a late arrival is one landing 15 minutes past its scheduled arrival time, airlines can increase their on-time percentages by increasing the scheduled flight time. See Whalen et al. (2007).

Jeffrey P. Cohen is an associate professor of economics at the Barney School of Business, University of Hartford. Cletus C. Coughlin is a vice president and deputy director of research and Lesli S. Ott is a senior research associate at the Federal Reserve Bank of St. Louis. Jeffrey P. Cohen acknowledges the support of the Lincoln Institute of Land Policy and the University of Hartford.

how a specific market-based approach, the use of auctions to determine the right to arrive or depart in a specific interval of time, might be used as part of a strategy to mitigate delays and congestion.

The focus on specific time intervals reflects the fact that the demand for passenger travel varies throughout a day. It is this variability that increases the odds that the demand for takeoff and landing slots at certain times during the day will exceed an airport's capacity. Takeoff and landing fees that do not vary throughout the course of a day are simply too low to efficiently allocate the scarce good of the right to land or depart during some time intervals. Armed with this economic insight and eager to solve the obvious congestion problem, many economists have suggested using a market-based approach.[3]

On September 16, 2008, the Federal Aviation Administration (FAA) announced its intention to implement a market-based approach by auctioning a limited number of slots at Newark, LaGuardia, and Kennedy airports starting on January 12, 2009.[4] Ideally, the auctions would ensure that the slots would be purchased by airlines placing the highest value on those slots. In addition, the FAA stated that the proceeds from the auctions would be spent on New York City area projects to mitigate congestion and delay.[5] Consequently, in a cost-effective way, the FAA hoped to achieve its goals: reduce some congestion directly and immediately and generate financing for additional projects to reduce congestion and delay in the future.

Despite the potential for the FAA's approach to alleviate delays, numerous parties voiced strong opposition against it—including the Air Transport Association of America (ATA), the Port Authority of New York and New Jersey (PANYNJ), and legislators.[6] In mid-December 2008 a three-judge panel of the U.S. Court of Appeals for the District of Columbia Circuit ruled that the slot auction could not be held until a federal court ruled on the objections raised by the ATA and the New York airport officials.[7] Effectively, the court decision shifted the question of whether to move forward with the auctions from the Bush administration to the Obama administration. Most recently, in spring 2009, the U.S. Department of Transportation announced the cancellation of the slot auctions.[8]

Despite the lack of immediate public-policy relevance, an examination of the FAA's proposal is still of much value because of the important economic issue that it addresses and the interesting economic and political arguments associated with it. Our primary goal here is to make the theory and the arguments, especially those relying on economics, accessible to those with little formal training in economics. We begin by identifying insights from economic theory that are relevant to dealing with congestion. Next, we summarize the FAA's proposal and then identify the key features of the proposed auction. With economic theory and the proposal as background, we then turn to the arguments for and against the FAA's proposal.

---

[3] See Whalen et al. (2007) for a discussion of failed policies that have attempted to reduce delays at LaGuardia, Kennedy, O'Hare, and Washington's Reagan National Airport.

[4] The actual use of auctions in the context of airports would be novel; however, auctions have been used in other settings, such as airwaves and pollution rights. As discussed by Tietenberg (2000), the approach can be viewed as a property-rights approach. A market is created by defining a property right and then allowing the right to be traded. Many issues immediately arise, one of which is who is given the right initially. A government agency might hold the right initially and then auction it off or it might allocate the right based on some history of activity or some other way. Concerning greenhouse gas emissions, the initial allocation is a key component of a cap-and-trade program. As discussed in *The Economist* (2009), the possible use of auctions for reducing emissions in the United States is controversial.

[5] Morrison and Winston (2008) provide suggestive evidence supporting such spending. They found that $1 of FAA spending reduced the costs of delay to airport users by $2.13 and that this spending could generate even larger benefits if it were allocated toward airports with the greatest delays.

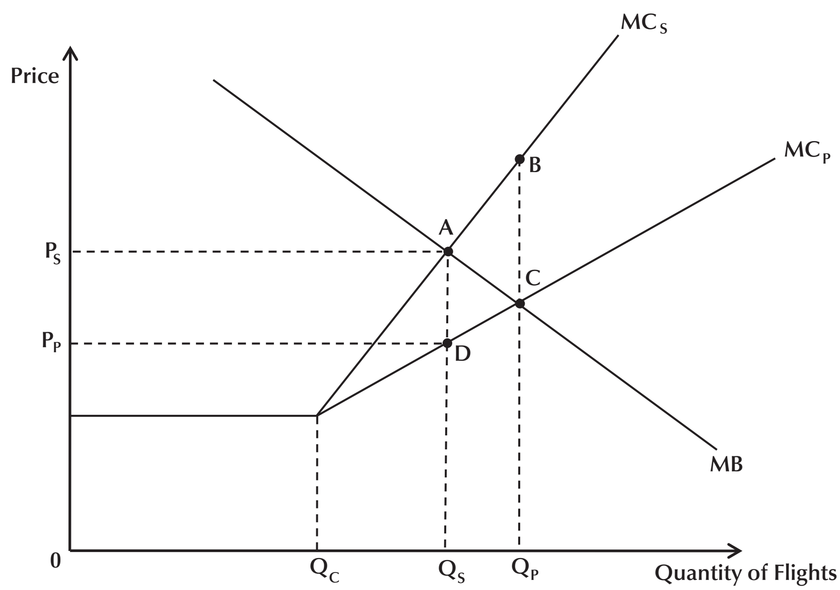
[6] The ATA, which represents the nation's largest airlines, filed a lawsuit to stop the auctions, arguing that the government lacked the legal authority to impose the auctions. The Port Authority of New York and New Jersey, which runs all three airports, supported the ATA. Senator Charles Schumer from New York characterized the proposed auctions as insanity and argued: "Auctions have never been tried and were hatched by a handful of ivory-tower types in the administration" (see Caterinicchia, 2008).

[7] See Holland (2008).

[8] See Bomkamp (2009).

Figure 1

The Efficient Quantity of Flights: Optimal Congestion



CONGESTION AT AIRPORTS: SOME BASIC ECONOMIC PRINCIPLES

Congestion problems arise at airports when the flight activity scheduled by more than one airline for a period of time cannot be accommodated in that time frame and delays occur. Congestion at airports is an example of what economists term a "negative externality." A negative externality occurs when an individual consumer or firm making a decision does not have to pay the full cost of the decision.[9] As a result, some costs are forced on other consumers and firms. The shifting of costs onto others means that social costs (i.e., the private costs plus the costs forced on others) exceed the private costs. When the decisionmaker does not bear the full costs, then the decisionmaker will engage in too much of the activity.

[9] Congestion is unlikely to be a problem at an airport dominated by one airline. Any congestion from that airline will delay primarily its own flights and, thus, the costs are borne by that airline. The airline will consider the effect of scheduling an additional flight on the revenues and expenses of its other existing flights, as well as the revenues and expenses of the additional flight.

Economic Efficiency and Optimal Congestion

For our illustration, we focus on the scheduling of flights by airlines.[10] When we use the term *scheduled*, we also assume *flown*. Thus, the costs and benefits identified below are for scheduled flights that occur. In Figure 1, the quantity of flights is on the horizontal axis, while a measure of the value (price) of a scheduled flight is on the vertical axis. Private marginal cost, MC_P, measures the cost of an additional flight. For a small number of flights, private marginal cost is drawn as a horizontal line indicating the same incremental cost of additional flights. Eventually, private marginal cost is positively sloped, suggesting that additional flights become increasingly costly. In other words, the costs borne by the scheduling airline increase after the number of flights reaches Q_C. Meanwhile, social marginal cost, MC_S, includes private marginal cost plus the cost forced on other airlines when a specific airline schedules an additional

[10] See Cohen and Coughlin (2003) for a similar discussion from the perspective of consumers.

flight. These latter costs are congestion costs, which are simply the difference between social and private marginal cost. In Figure 1, for a small number of flights the social and private marginal cost curves coincide: This reflects an absence of congestion costs. As the number of flights increases beyond Q_C, however, congestion occurs and social marginal cost lies above private marginal cost.

Airlines schedule flights to maximize their profits. The marginal benefit curve, which ultimately hinges on consumer demand, reflects the benefits (i.e., revenues) that the airline generates from a flight. These additional benefits decline as the number of scheduled flights increases.[11] As a result, the number of scheduled flights in an unregulated, competitive market will be Q_P.[12]

From society's point of view, Q_P is an excessive number of flights because at this quantity of flights social marginal cost exceeds marginal benefit. Ideally, Q_S should be the number of scheduled flights because, for a quantity of flights less than Q_S, the marginal benefits for scheduling additional flights exceed marginal social cost. Beyond Q_S, marginal benefits are less than marginal social costs. Note that some congestion exists at Q_S, so the optimal level of congestion is not zero, but rather some positive amount. The issue for policymakers is how to reduce the number of flights from Q_P to Q_S.[13]

One option is to allow no more than a given number of landings and departures in a specific period.[14] In Figure 1, this means reducing the quantity of flights to Q_S.[15] Quantity-based regulation would limit the number of flights to this level. An important issue involves deciding who is allowed to use the scarce arrival and departure slots. This can be done by maintaining the existing flight shares of airlines before the reductions, but this hinders potential new entrants and might cause some airlines to retain slots that they value less than other airlines might value them. Also, because the slots are valuable, the use of this option means that the government generates none of the potential revenue. Securing such revenue is part of a process termed "value capture." Forgoing this revenue might make the financing of future airport expansions reliant on less-efficient options, such as other taxes. See the boxed insert for a detailed discussion of the revenue-raising technique of value capture.

To overcome the likely inefficiency of simply continuing the existing flight shares, two market-based measures of allocating slots have been proposed, one relying on a price-setting mechanism and the other on a quantity-setting mechanism. Both measures can yield identical results, but considerations, such as uncertainty involving demand or supply and the cost of implementation, might lead to the superiority of one measure.

A price-setting mechanism, known as "congestion pricing," varies landing fees by time of day. With this approach regulators would set prices for landings and takeoffs that would yield the efficient level of output at the airport. In light of the excess demand for arrival and departure slots at certain times of the day, the goal of congestion pricing is to shift some demand during the most congested periods of the day to times when capacity is readily available (i.e., periods of excess supply). Rather than have fees that do not vary over the course of a day, access fees would be higher during peak travel hours to induce airlines to shift some operations from the peak travel hours to nonpeak travel hours. In Figure 1, a congestion tax equal to AD (or the difference between P_S and P_P) per flight would induce airlines to reduce scheduled flights from Q_P to Q_S.

Auctions are another market-based mechanism to mitigate the adverse effects of congestion and

[11] This is analogous to the fact that lower airline fares are associated with an increase in quantity demanded on the part of airline passengers.

[12] The welfare loss stemming from the excessive number of flights is the area ABC in Figure 1.

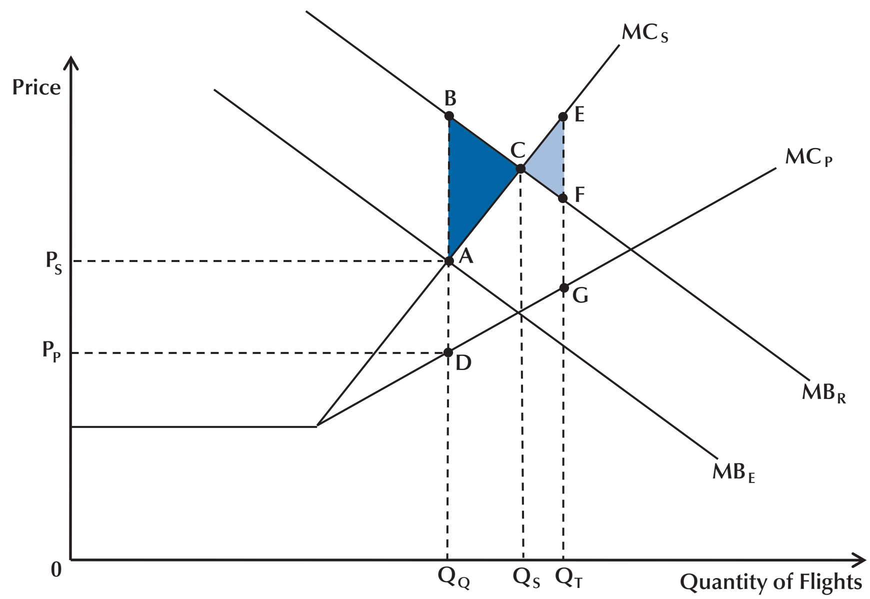
[13] Congestion *might* be reduced by increasing capacity via adding runways and improving the air traffic control system, but these approaches do not deal with the externality problem. See Cohen and Coughlin (2003) for an introduction to the economics of airport expansions. A counterintuitive proposal for dealing with congestion has been highlighted by Dubner (2009). Closing LaGuardia would free up air space and allow Newark and Kennedy to operate more efficiently and at a higher capacity.

[14] Safety considerations also affect the actual number of arrivals and departures that an airport can handle.

[15] A similar option mentioned in Whalen et al. (2007) is to allow the airlines to negotiate with each other to reach a solution. Such an option is likely flawed because the interests of potential competitors and consumers would not be adequately served.

Figure 2

Prices Versus Quantity Regulation: Marginal Benefit Uncertainty



allocate scarce arrival and departure capacity efficiently. Auctions are viewed as a quantity-setting mechanism. Under this option, property rights to use the Q_S slots are sold or leased to the highest bidders via an auction. The winning bidders would have the right to use the slots for some specific time period each day for the length of the contract or sell or lease the right in a secondary market.[16] With a second price auction, airlines would be induced to bid their true valuation for the slots, and the resulting payment would extract nearly the entire "surplus" from the airlines.[17] Thus, the auctions lead to the efficient outcome.[18]

[16] Another feature of the property right, recommended by Whalen et al. (2007), is to include a cancellation priority. This would provide a way to allocate the slots when weather or some other factor temporarily reduced the number of arrivals and takeoffs that could be handled.

[17] In a second price auction, the highest bidder wins but pays a price equal to the second-highest bid.

[18] The economic roots of our discussion can be found in a classic paper by Weitzman (1974) that compares environmental regulation using a price (e.g., a tax on pollution emissions) with regulation using a quantity (e.g., emission standards).

Adding Real-World Complexity

In the preceding example, the regulators were assumed to know the locations of the marginal benefit and marginal cost curves with certainty. A more realistic assumption is one of uncertainty, which allows for the regulators' expectations to differ from what actually occurs. This modification can cause the results of price regulation to differ from quantity regulation. Assume that the regulators are certain about the locations of the marginal cost curves but are uncertain about the location of the marginal benefit curve.[19] In Figure 2 this is represented by a realized marginal benefit (MB_R) curve that lies above the expected marginal benefit (MB_E) curve.[20]

[19] This example can be characterized as illustrating regulatory cost uncertainty. Albeit confusing at first, the reasoning is straightforward. The benefits of regulation are the benefits of reducing congestion costs, while the costs of regulation are primarily the foregone travel benefits. Thus, in the example, the regulators are certain about the private and social (i.e., congestion) costs of airline flights, but are uncertain about the benefits of flights.

[20] See Adar and Griffin (1976) for a discussion of uncertainty with linear marginal benefit and cost functions.

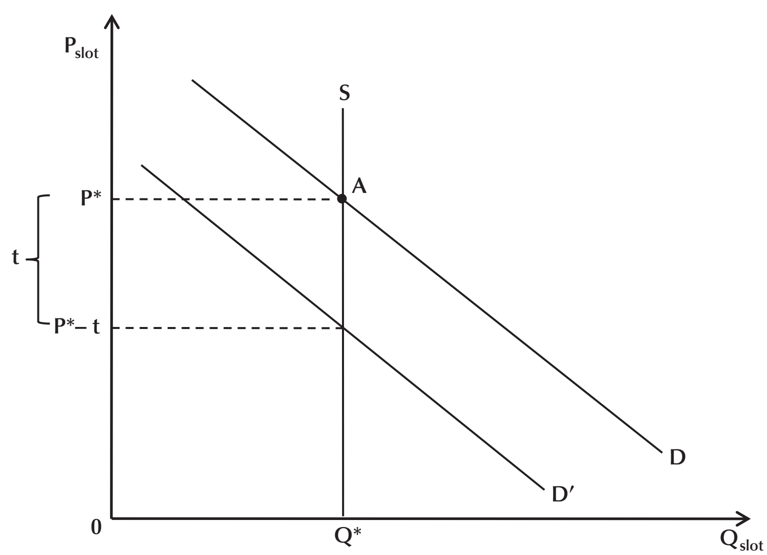
VALUE CAPTURE: ECONOMIC THEORY

Generally speaking, the use of most types of taxes to fund public projects creates some economic inefficiency. For example, a sales tax puts a wedge between the price paid by a buyer and the price received by a seller. Relative to a situation without a tax (i.e., the price is the same regardless of whether you are a buyer or a seller), imposing a tax tends to cause the price paid by buyers to increase and the price received by sellers to decrease. As a result, some mutually beneficial exchanges do not occur. Economists refer to the forgone benefits as a "deadweight loss."

This general result, however, does not hold when the supply curve is fixed at a specific quantity (that is, when supply is perfectly inelastic). Capping the number of slots would produce such a supply curve. In the accompanying figure, with the price of slots on the vertical axis (P_{slot}), and the quantity of slots (Q_{slot}) on the horizontal axis, the supply of slots is drawn as a vertical line at Q*. In other words, suppose the airport (or the FAA) sets and controls the initial supply of "slots" by imposing operating limits and distributes these "slots" free of charge to the airlines. As discussed in the text, the determination of the socially ideal level of slots is far from easy. For the remainder of this discussion, we assume that Q* is the ideal level of slots. Meanwhile, the demand for slots is drawn with a negative slope, reflecting that the quantity demanded of this input for flights increases as its price declines. In other words, airlines will prefer to increase their slots as the price of slots decreases.

Figure B1

Taxing Perfectly Inelastically Supplied Slots



Because the supply of slots is perfectly inelastic, a tax (t) imposed on the owners of these slots will extract the maximum amount of tax revenue (or value) from the slot holders without distorting behavior of the "consumers" of slots.[1,2] In other words, there is no deadweight loss associated with this tax. The tax on slots would shift the demand curve downward, from D to D′, by the amount of the tax. The after-tax price received by sellers of slots would be ($P^* - t$). Tax revenues would be the amount given by $t \times Q^*$.

To maximize revenue, the airport authorities would want to tax until tax revenues are slightly less than the area of the rectangle P^*0Q^*A. This scenario implies that the after-tax value of a slot is only slightly greater than zero, while virtually the entire surplus is captured by the taxing authority. The tax revenues could be used to finance airport operations, maintenance, and/or long-run airport improvements, and this approach would avoid the need for higher other distortionary taxes.

One might ask the following question: Instead of going through these steps to generate the highest possible amount of tax revenues by extracting the entire surplus, why not have the government sell these slots or hold a first-price auction?

There are two reasons. First, a practical problem arises in determining beforehand the position of the demand curve at different slot prices. In the typical case, the demand curve represents the airlines' reservation price. In the present case, it is not feasible to ask the airlines what their reservation price is for a given slot, which is the approach of a first-price auction, because of the incentive for airlines to understate their true valuation. A second-price auction is a plausible alternative because it is known to induce participants to bid their true valuation. In other words, the second-price auction organized by the government would be expected to extract the same amount of revenue from the airlines as a value capture tax on the slot holders in the amount of P^*. So a second-price auction can be used as an equally effective (and efficient) method of value capture.

Second, to auction the slots, the government must rescind operating rights and possibly compensate the current occupants of the slots. We have shown that the government can generate the same revenues from a second-price auction as from a value capture tax. However, due to political pressures, the government likely would somehow need to compensate airlines currently holding the slots after reselling them at auction. As a result, net of these compensation costs, the government may actually end up with less revenue if it auctions slots than if it taxes current slot holders. Thus, from a political perspective, a value capture tax on the current slot holders may be a better approach to raising revenues than an auction, although the value capture tax may not solve the congestion problem in the same way as an auction. But, if the government were to reset the maximum numbers of flights and take back only some of the existing slots and impose a value capture tax on the remaining slots, this could be at least as effective as auctioning all of the slots in addressing congestion. It may even raise more revenues than an auction because fewer flights would have their slots confiscated and in turn, there would be less need for compensation.

---

1 See Cohen and Coughlin (2005) for an exposition of the optimal taxation of land, which they assume is also perfectly inelastically supplied.

2 Similar to landing slot taxation, Tietenberg (2000) discusses the evolution of the current tradable discharge permits system in the United States. He notes that initially ozone depletion permits were given to pollution-generating firms and, later, these firms were taxed by the federal government because of the rents generated from the trading of these permits and the desire by Congress to generate revenues.

The regulators make their decision based on expectations, so they would impose a tax of AD per flight if regulating by price and a maximum number of flights of Q_Q if regulating by quantity. Given a tax of AD per flight, the number of scheduled flights would be Q_T. The underlying reasoning is as follows: From the perspective of airlines, the tax of AD causes a parallel upward shift of their costs. This line, which is not drawn, is the private marginal cost curve plus the tax of AD. The intersection of this line with the realized marginal benefit determines the quantity of scheduled flights, which as stated previously is Q_T. Note that in Figure 2 the length of AD must equal FG. In this case, the quantity of flights would exceed the efficient quantity, Q_S, which is determined by the intersection of the social marginal cost curve and the realized marginal benefit curve. On the other hand, regulation by quantity would lead to too few flights because Q_Q would be less than Q_S. Thus, neither form of regulation is efficient.

The welfare loss, measured as the deviation from economic efficiency, associated with price regulation is represented by the area of triangle CEF, while the welfare loss associated with quantity regulation is represented by the area of triangle ABC.[21] The preferred regulatory instrument is the one yielding the smaller welfare loss. As drawn in Figure 2, the area of triangle CEF is smaller than the area of triangle ABC, so price regulation is preferred to quantity regulation. However, we could have drawn Figure 2 so that quantity regulation is preferred to price regulation. Assuming straight lines, the flatter the slope of the realized marginal benefit curve relative to the slope of the social marginal cost curve, the more likely quantity regulation will become the preferred approach.[22]

The preceding example is focused on uncertainty involving the marginal benefit curve. What happens when the uncertainty is restricted to social marginal costs? The answer is that price regulation and quantity regulation generate identical results—regardless of the slopes of the curves.[23]

The preceding discussion ignores the possibility that congestion at an airport depends on congestion at other airports. Two types of interdependencies—substitutability and complementarity—exist. In the first case, nearby airports compete with each other for passengers and cargo, so that increased traffic at one airport might reduce the traffic at a nearby airport. On the other hand, due to the network character of the air transportation system, airports also provide complementary services because a takeoff from one airport requires a landing at another airport. Thus, the increased use of runway slots at one airport tends to increase the demand for runway slots at other airports.

Czerny (2006) focuses on demand complementarity and compares the welfare effects of quantity regulation via slot constraints with price regulation via congestion pricing. His key conclusion is that the demand-related features of the airline industry increase the attractiveness of using slot constraints relative to congestion pricing. With congestion pricing airport usage is uncertain, and demand complementarity causes this demand uncertainty to propagate from one airport to another. Slot constraints can eliminate this propagation and can prevent the excessive use of runways and, thus, generate a preferred solution.

In contrast to Czerny (2006), Brueckner (2009) ignores both demand uncertainty and network externalities. He focuses on one airport served by more than one airline and allows the number of flights to differ across airlines. In his model, an individual airline accounts for the congestion costs that it imposes on itself (i.e., the airline internalizes congestion). He explores two price-based regimes and two quantity-based regimes.

Under a price-based regime, the airport authority announces a charge per flight for airlines to use a congested airport. The airlines then decide

[21] The welfare loss associated with price regulation can be viewed in the following way: For each flight in excess of Q_S, marginal costs exceed marginal benefits. The triangle CEF measures the net total cost of these excessive flights. For quantity regulation, too few flights are scheduled because at Q_Q marginal benefits exceed marginal costs. The triangle ABC measures the net total benefits that are forgone by not having Q_S flights.

[22] This general result is not altered when the realized marginal cost curve lies to the right of the expected marginal cost curve.

[23] An assumption underlying the preceding analysis is that the uncertain benefit and costs functions are independent. Assuming a positive (negative) correlation of benefits and costs, Stavins (1996) finds that regulation by quantity (price) is likely preferred to regulation by price (quantity).

the number of flights. Two approaches are possible for the setting of prices. One approach entails charges that can differ across airlines, while the second approach entails an identical charge for each flight regardless of the airline. Under the first approach, assuming slot charges are set correctly, it is possible to produce the social optimum. Prices must vary across airlines because of the combination of different sizes of airlines with the fact that an airline takes into consideration the congestion it imposes on itself. Moreover, assuming two carriers, the larger carrier pays a lower congestion charge than the smaller carrier.[24]

Turning to the quantity-based regimes, the airport authority must begin by announcing a fixed volume of flights. The fixed (i.e., socially optimal) number of slots can be allocated either by (i) distributing the slots free of charge and allowing carriers to make adjustments through trading or (ii) auctioning the slots. Regardless of the allocation mechanism, the quantity-based regimes can produce a socially optimal result. This result is an illustration of the Coase (1960) theorem. When trade in an externality is possible and no (or sufficiently low) transaction costs exist, then an efficient outcome will occur regardless of the initial allocation of property rights.[25]

In sum, the scarcity associated with slot capacity can be addressed by having airplanes and travelers wait, politically deciding winners and losers, or by pricing the scarcity via a market mechanism. Economists tend to prefer the third option because of its efficiency properties. Relative to government decisions, the market mechanism allows for a speedier response to market changes. Moreover, the results based on economic theory applied to airport usage provide a justification for using quantity regulation for runway usage. We now turn to the details of the FAA's plan and then the objections that have been raised.

DETAILS OF THE FAA'S CONGESTION MANAGEMENT PLAN

On October 18, 2008, the FAA published its final rules concerning congestion management for LaGuardia, Kennedy, and Newark airports.[26] The FAA identified two primary methods to alleviate congestion. First, the current cap of 81 on the number of hourly takeoff and landing slot operations available at Kennedy and Newark airports during peak hours was extended and the number of peak hourly slot operations allowed at LaGuardia was reduced from 75 to 71.[27,28] Second, five consecutive annual slot auctions for a small number of slots at each airport (approximately 2 percent per year) would occur so that, in the face of the hourly caps on arrivals/departures, individual air carriers could attempt to increase their slot holdings through competitive bidding.[29] The funds collected from the slot auctions would be used to further mitigate congestion in the New York City area.[30]

The reduction in the total number of slots available at LaGuardia would have required a slot reallocation among operating airlines. According to its final rule, the FAA would have determined the structure of this reallocation. Viewing the slot reallocation at LaGuardia as integral for congestion management, the FAA would have

[24] This result occurs because the larger carrier (i.e., the one with more flights) internalizes more of the congestion damage from its operation of an additional flight than does the smaller carrier. In this case, the common charge will penalize the larger (smaller) carrier too much (little) for the congestion it causes, resulting in the flight volume for the larger (smaller) carrier being too small (large).

[25] Transaction costs as well as strategic behavior can prevent the parties involved from reaching an agreement that is efficient. In addition, there are other considerations that could produce a preference for one allocation method over the other. For those airlines allocated slots free of charge, the receipt of such a valuable asset free of charge would be especially attractive. On the other hand, the airport authority would be forgoing revenue that could be used to improve the airport and the air transportation system.

[26] The final rules can be found at www.fly.faa.gov/ecvrs/circulars/FR_JFK_and_EWR_final_rule_101008.pdf and www.fly.faa.gov/ecvrs/circulars/FR_LGA_Final_Rule_101008.pdf.

[27] The caps at Kennedy and Newark were originally imposed in the first half of 2008. Operational caps at all three airports are in effect from 6 a.m. until 9:59 p.m. each day.

[28] The FAA estimated that the new operational cap at LaGuardia would improve flight delays there by 41 percent. See page 60574 of the final rule for La Guardia (*Federal Register*, 2008b).

[29] A "slot," as defined by the FAA, is the right to land or depart during a 30-minute window.

[30] The FAA estimated the amount of net benefit from the LaGuardia slot auction would be $65.4 million. See page 60595 of the final rule for LaGuardia (*Federal Register*, 2008b).

required the same structured slot reallocation at Kennedy and Newark airports as well.

The terms of the slot reallocations were as follows: The FAA planned to grandfather 85 percent of total slots to existing slot holders, permanently retire 5 percent, and award the last 10 percent in increments over 5 consecutive years to the winners of annual slot auctions. The FAA would have allowed each air carrier to choose half of the 15 percent of current slots they were to surrender. The FAA would have decided the remainder to ensure that slots from every peak hour of the day were available both for permanent retirement and annual auction. In exchange for the terms of the slot reallocations, the FAA would have granted to each airline 10-year slot lease rights to all slots in its new allocation.[31]

Though the FAA intended to hold only one slot auction per year, it recognized that increased competition coupled with the operational caps might leave some airlines with fewer slots than desired after the first slot auction had taken place.[32,33] Therefore, a secondary market by which air carriers could buy and sell slots among themselves, subject to FAA oversight, was to be encouraged after the first slot auction was completed. Secondary market operations were to be allowed not only for slots won via the auction process but for all slots.[34,35]

The FAA's intention was that the secondary market for slots would grow and be sufficient to ensure an efficient allocation of slots long after the current congestion management rule expired in 2019. In other words, each slot would be obtained by the air carrier that valued it most. For some, maximizing a slot's potential would have meant using larger planes to transport more passengers at a given time. For others, it would have meant subleasing the slot if the gains from the sublease would have been greater than the marginal benefit from using it for their own business operations. In both scenarios, congestion would have been mitigated and travelers would have benefitted.

DETAILS OF THE DESIGN OF THE FIRST FAA SLOT AUCTION

According to the final rules, the FAA's first slot auction would have offered the lease rights to at least 24 takeoff and landing slots at LaGuardia and approximately 18 slots each at Kennedy and Newark.[36,37] While these slot auctions were considered by the FAA to be integral to their long-term congestion management plans, the overall design of the slot auctions was a critical factor for success.

The design of any successful commercial auction must incorporate the following three features: (i) Its format must be well suited to the characteristics of the goods being auctioned; (ii) it must allow winners to be determined both quickly and fairly; and (iii) it must generate final prices that reflect a good's true economic value. We will now examine how the FAA's proposed slot auction design addressed these concerns.

Combinatorial Auctions

That an even number of slots at each airport would have been auctioned is not by chance.[38]

[31] Though some disagree, the FAA states that up to this point no carrier has held a proprietary right to any slot, and that they have held the right to withdraw slots from carriers at any time. See page 60574 of the final rule for LaGuardia (*Federal Register*, 2008b).

[32] As noted earlier, the FAA would have held five annual slot auctions. This is the number of years it deemed sufficient for firmly establishing slot prices given its decision to hold only one auction per year.

[33] It is worth noting that, in some cases, the first auction may have actually left some airlines with more slots than desired given the newly established market values associated with them.

[34] The FAA would also have allowed any air carrier to present slots for auction during any one of the five official slot auctions. The airline could have set a minimum price for each slot and retained all auction proceeds.

[35] The FAA planned to provide ongoing support for this secondary market. It would have allowed air carriers to post available and desired slot subleases on its website, and it would have made slot transactions transparent by posting their values after a sale. It would also have monitored the secondary market for any anticompetitive behavior. Each airline would have retained all monetary gains from these transactions.

[36] See page 60547 of the FAA's final rule for Kennedy and Newark (*Federal Register*, 2008a).

[37] The four subsequent auctions were to offer at least 22 slots at LaGuardia and at least 18 at Kennedy and Newark.

[38] That the overall number of slots to be auctioned was small was also not by chance. The forced reallocation of a relatively small number of flights was also designed to cause as little disturbance as possible to overall airport/airline operations.

Takeoff and landing slots are seen as complementary goods by most would-be auction participants. Understandably, in most circumstances an airline would want to own the right to a takeoff slot if and only if it also had the right to a subsequent landing slot at that same airport (i.e., it needs the ability to conduct a round trip). Therefore, if bidders could bid for multiple slots, they would be more likely to express their true valuation for them—as some slots would be valued more as one of a pair. Consequently, the FAA had determined that the first slot auction would have been conducted as a combinatorial auction where participants may bid on combinations of slots rather than being restricted to bidding for single slots only.[39]

Had the auction been designed as strictly for single items, air carriers likely would have incorporated into their bidding strategy the possibility of a worse situation should they win only one slot when a pair was needed.[40] One can easily see how this could lead to distorted bidding for slots to hedge against that outcome. Moreover, a distorted valuation of slots in the primary market would likely lead to price distortion, and therefore inefficiencies, in the larger secondary slot market as well—an outcome the FAA wanted to avoid.

Combinatorial auctions are not without their drawbacks, however. While theory has shown this auction type as the best way of auctioning complementary goods, real-world combinatorial auctions are rare. Even with modern computing, determining the winner of a combinatorial auction within a realistic amount of time can be difficult. As we will see, the mathematics underlying combinations is the culprit.

Combinatorial auctions are desirable specifically because participants may fully express their preferences for any desired combination of goods, but so too are they problematic for the same reason. For any m items available in a combinatorial auction, there are 2^m-1 possible non-empty bidding combinations. As one can imagine, offering all but a small number of items for auction would lead to extreme difficulty for the seller in evaluating each bidding combination.[41] For example, the 24 slots offered in the first LaGuardia slot auction would have theoretically allowed for 16.8 million subsets of bids per bidder to be placed and, thus, to be evaluated. Even using today's computers the evaluation of 16.8 million subsets of bids would prove extremely time-consuming. Finding the solution for a combinatorial auction in which 2^m-1 bids may be submitted is known, in computational parlance, as an exponential time problem.[42] Given the FAA's desire to determine winners within 2 days of the auction's end, allowing bidders to submit 16.8 million sets of bids would have, in fact, made winner determination impossible.[43]

For a real-world combinatorial auction, this winner determination problem requires the auction to be designed such that a trade-off between theoretical winner optimization and time is made. To do this, restrictions on bidding must be in place. One approach is to require that bids be made via a particular bidding language. Another is to set a limit on the number of bids allowed. The FAA would have done both.[44]

A "bidding language" is a computer-friendly language that allows bidders to more succinctly (computationally speaking) represent their desired combinations of items. Slot auction participants would have likely used a computer program into which bidders would have input their slot preferences. That program would have then translated those preferences into the predetermined bidding language. As such, all bids would have been represented in a way that a computer algorithm could have more easily selected the winning combination.

[39] The auction would also have been sealed-bid and single round.

[40] In auction theory, this is known as the exposure problem.

[41] Likewise, a bidder can be overwhelmed by the number of bids he may submit and, therefore, be less likely to participate.

[42] Exponential time problems where m is sufficiently large are contained in a class of mathematical problems formally defined as NP-complete. For a broader discussion of the limitations of modern computers with respect to NP-complete problems, see Aaronson (2008).

[43] See slide 58 of the "2009 New York Slot Auctions Bidders Seminar" (FAA, 2008b).

[44] For an excellent discussion of the real-world complexities of combinatorial auctions, including other approaches to the winner determination problem, see Cramton et al. (2006).

Figure 3

Five Bidders for Two Slots

- b_1 {A} = 28
- b_2 {B} = 20
- b_3 {AB} = 32
- b_4 {A} = 14
- b_5 {B} = 12

Winners

Vickrey Prices:
$p_1 = 14$
$p_2 = 12$

To further increase efficiency, the FAA planned to restrict the maximum number of allowable bids to 2,000 per auction participant. While 2,000 is a small fraction of the theoretically possible 16.8 million, the FAA had determined that this number of bids would amply allow each carrier to express all preferences and, therefore, would not have compromised the fairness of the results.[45] Once all bids were received, the winners of each slot auction were to be identified by determining the collection of bids that maximized total revenue to the FAA subject to the constraints that no participant could win more than one combination of slots per airport and that no slot could be awarded more than once.[46]

It is reasonable to assume that, when notified, winners would be expected to remit payment equal to that of their winning bid. However, for the slot auctions, this would not have necessarily been the case. To deter bidding distortions, the FAA would have required winners to make a final payment consistent only with the bids of their strongest competitors.[47] This so-called bidder-optimal core pricing strategy is similar, in auction theory, to a Vickrey second-price auction where the winner of a single good pays the price of the second-highest bid for that good.[48] However, given that the goods in the FAA slot auction could have been bid for in multiple combinations, in some cases the Vickrey price may have been too low.

Determining Bidder-Optimal Core Prices

The following figures will demonstrate how, in the case of multiple bidding, final bidder-optimal core prices were to be determined.[49] Figure 3 shows the bids of five bidders (b_1, b_2, b_3, b_4, and b_5) competing for the lease rights to two slots (A and B). Subject to the constraints that no slot may be awarded more than once and that total revenue should be maximized, bidders 1 and 2 are declared the winners with a total bid of 48 for the two slots.

The Vickrey price for slot A is 14 while the Vickrey price for slot B is 12. However, bidder 3's bid of 32 for both A and B together, while still smaller than the total for the winning bids, exceeds the 26 that would be realized if winners were made to pay individual second prices. Therefore, in terms of a second-price pricing strategy, the seller will require a payment of at least 32 before awarding slots A and B. This is because the price of 32 for the joint A and B slots is the second-highest price compared with the price bidders 1 and 2 would bid if bidders 1 and 2 were to bid jointly on slots A and B.

Figure 4 identifies the competitive pricing core based on the five original bids. This competitive core area represents all prices between the revenue maximizing outcome of 48 and the next-best bid of 32. The competitive pricing core consists of all feasible combinations of prices that yield an efficient allocation of the two slots. In the present case, efficiency is directly related to the willingness to pay by the bidders. Ultimately,

[45] Interestingly, this bid limitation may have actually *increased* auction participation by carriers as the total number of possible bids is more manageable even though some may argue that it is theoretically limiting.

[46] See Amendment 1 of the auction procedures (FAA, 2008a).

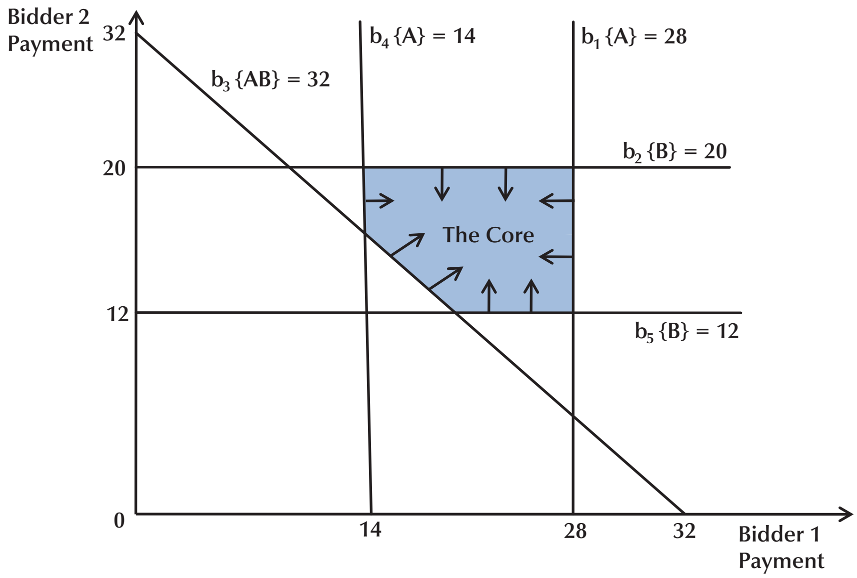
[47] Should a tie have occurred, it would have been randomly broken and the declared winner would have paid the full amount of his or her bid. Also, should only one carrier have bid for a particular slot, the competitive price for that slot would have been a nominal reserve price predetermined by the FAA.

[48] The following discussion relies on core theory. The core can be viewed as the set of feasible allocations that cannot be improved upon by trades within a subset of the economy's consumers. See Telser (1994).

[49] The figures, with some modifications, are from slides 25 through 29 of the auction bidders' seminar (see FAA, 2008b).

Figure 4

The Competitive Pricing Core



their willingness to pay ensures that the final prices reflect the economic value of the slots.

Given that bidders 1 and 2 are the winners based on their willingness to pay for the slots, the competitive pricing core consists of all feasible combinations of prices that yield at least 32. Because bidder 1 is willing to pay up to 28 for slot A, then any price above 28 is not feasible for bidder 1. In addition, bidder 1 must pay at least 14 because bidder 4 was willing to pay 14. Thus, the portions of the graph to the right of the vertical line at 28 and to the left of the vertical line at 14 are eliminated. Similarly, any price above 20 and below 12 is not feasible for bidder 2, which eliminates the portions of the graph above the horizontal line at 20 and below the horizontal line at 12. Finally, the competitive pricing core requires that the winning bidders jointly pay at least 32. The diagonal line provides the fifth and final side of the competitive pricing core.

To identify the final price each winning bidder may pay, the seller will reduce both bidder 1 and 2's bids subject to the constraint that jointly they must pay at least 32. This range of bidder payments is identified in Figure 5 as the portion of the diagonal line (i.e., the border of the competitive pricing core) labeled the bidder-optimal core. The bidder-optimal core is all points of intersection between bidder 3's bid of 32 and bidder 1 and 2's possible final individual payments. However, given that there is no unique bidder-optimal core solution, how does the seller settle on individual final prices that are fair to both bidder 1 and bidder 2?

The seller chooses the bidder-optimal core price closest to the total Vickrey price of 26. As seen in Figure 6, this is the point on the graph representing the shortest distance between the Vickrey prices and the bidder-optimal core. According to Figure 6, bidder 1 would pay a final price of 17 while bidder 2 would pay a final price of 15.

Our elementary discussion suggests that the use of a combinatorial auction is far from straightforward. One of the many issues for policymakers is whether the likely benefits of such an auction in terms of economic efficiency and revenues for congestion-related projects outweigh its likely

Figure 5

Bidder-Optimal Core Prices

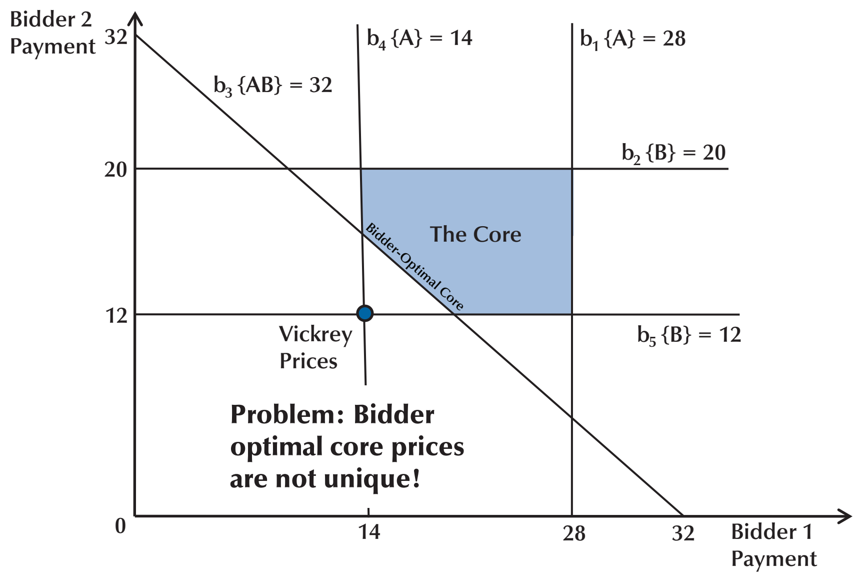
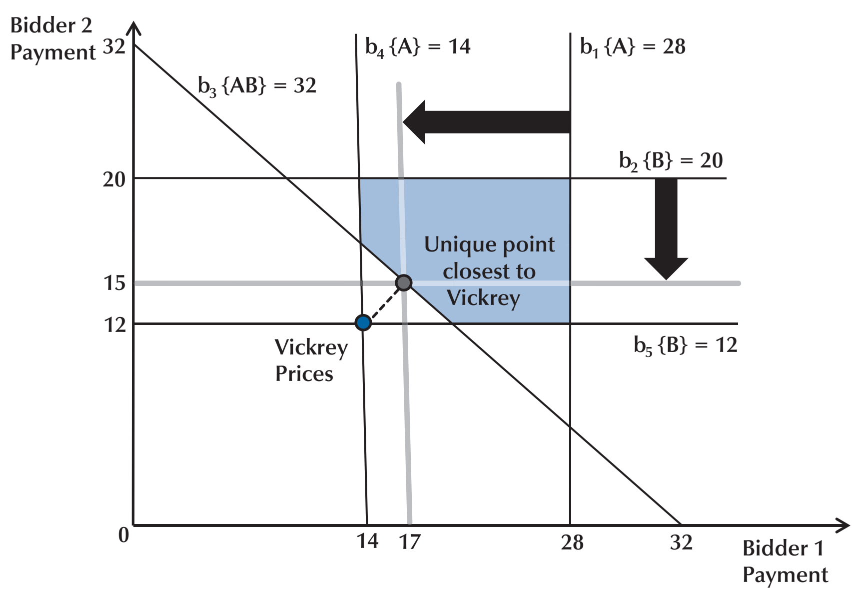


Figure 6

Winning Bids



costs. We now turn our attention to additional issues raised by the FAA's proposal.

A DISCUSSION OF ISSUES CONCERNING THE PROPOSAL

Numerous issues exist with using any market-based mechanism to deal with congestion. Our discussion is focused on economic issues, so we do not examine whether the FAA has the legal authority for its proposed rule.[50] Many objections highlight the harm that would be done to some group. Given that the implementation of any market-based option would cause some prices and airline transportation services to change, it is reasonable to expect harm for some groups. Some concerns stress the harm that would result if the auction plan were not implemented properly, while others stress the harm that some would bear even with proper plan implementation because they perceive certain features of the plan as unnecessary and counterproductive.

To date, auctions and congestion pricing have not been implemented on a broad scale at any U.S. airport. Many worry that prior experience with auctions and congestion pricing in other situations is not especially useful because of the uniqueness and complexity of the airline industry and its networks. Given the fragile financial situations of many airlines, the implementation of an ill-designed scheme could be harmful to travelers, airlines, and other stakeholders, such as the New York City region. One specific area of concern for such harm relates to the fact that both the airlines and the Port Authority have much debt and other financing tied to service levels.[51] Changes to the service levels could lead to economic trauma.[52]

The preceding discussion leads naturally to the fact that the implementation of a congestion pricing or auction mechanism would entail transition costs. Obviously, the extent of these costs depends on the actual mechanism selected. The new situation could have a major impact on scheduling practices and require new investment to manage the new mechanism. Nonetheless, it is expected that these costs would prove minor and be offset over time by the gains in economic efficiency.

Given that the decision to use auctions is a political as well as an economic decision, the following question immediately comes to mind: Would there be so many exemptions to deal with these concerns that the use of an auction would do little to increase economic efficiency? Increased exemptions means higher prices for those not receiving exemptions and quite possibly those travelers would see few benefits in the form of operational efficiency or overall system performance.

The existing system does not generate an efficient use of resources and, as a result, there is overuse and excessive costs during peak hours. While the partial elimination of service is a possibility, air carriers have other possibilities that might be viable, such as flying at different (non-peak) times and using other airports, such as reliever airports for some flights.[53] A key point involving auctions is that the market mechanism provides incentives to use existing resources more efficiently.

---

[50] Many ideas in this section were identified by Working Group 2 in *New York Aviation Rulemaking Committee Report*, December 13, 2007. Also, for a more detailed discussion of institutional issues related to slot auctions, see Levine (2009).

[51] A related issue concerns the "harm" for certain travelers. Without question, the use of auction pricing would cause some price changes and service changes. The monetary cost to travelers during the peak hours in the New York City area would likely rise and reduce general aviation access at those times; however, the nonmonetary costs of delays and congestion should decline. Some travelers would likely be harmed by higher airfares and reduced flight options, particularly those flying to small communities. Mid-size markets might also lose service. These markets demand high-frequency service that is provided by regional jets. Without exemptions from congestion fees or auction prices, service to some or all of these communities would be reduced. But Whalen et al. (2007) stress that the elimination of what are termed "carve-outs" for corporate jets and general aviation users is a key aspect of their auction proposal.

[52] A poorly implemented plan could damage the industry. In a comment on the proposal to auction slots at LaGuardia Airport, the Federal Trade Commission (2008) specifically highlighted auctions involving electricity markets in California as a fundamental factor in the market meltdown. The issue is the likelihood of a poorly designed and implemented plan. This is possible; however, if done well, then the reduction of congestion costs should increase system reliability and provide savings related to time costs for consumers.

[53] In the present case, PANYNJ has purchased a lease for Stewart Airport (in New Windsor, New York, 60 miles north of New York City) as a reliever airport.

While auctions focus on the efficient use of slots, the allocation of slots also has important implications for the use of gates, hangars, and other physical assets. An airline that loses slots might incur losses due to declines in value on these other assets. Moreover, there is no clear mechanism whereby the winners of the slots can acquire access to these other complementary assets controlled by the losers. A longer-run possibility is that an airline might be less willing to make or support infrastructure investments in the future. With the advent of auctions and congestion pricing, the New York market is likely to be viewed as riskier and thus investment by airlines might be deterred. This complexity might be difficult to resolve without a higher governmental authority; yet, there would be incentives to reach mutually beneficial agreements. A secondary market for these other assets would likely resolve some of these concerns.[54,55]

Another set of concerns revolves around the fact that the U.S. airline industry operates in an international environment and is subject to international rules. For example, international routes are heavily dependent on connecting flights for domestic passengers, so anything that affects the connecting flights (times, frequencies, and so on) might make the related international flight unprofitable.[56]

A second international-related concern is whether foreign carriers would also be subject to congestion pricing or auctions. On efficiency grounds, all carriers should be subject to the same rules. As suggested previously in the context of an exemption granted to a specific U.S. carrier, any exemption of foreign carriers would put U.S. carriers at a disadvantage.[57]

Lastly, auction and congestion pricing schemes might violate U.S. bilateral and multilateral agreements, such as the U.S.–European Union Open Skies Agreement, because such charges are not cost based. The charges might be viewed as exceeding the full cost of providing the relevant air traffic control or airport services. From an economist's view, congestion is a true cost. Thus, from a theoretical perspective of an economist the schemes might not violate the law, but that does not mean the courts would agree.

So far, most discussion in this section has focused on winners and losers. In their comment on the proposal regarding LaGuardia, the ATA and their consultant argued for modifying the proposal in a way they felt would remove both complexity and uncertainty while retaining the benefits.[58] They argued that the proposed rule for LaGuardia could be separated into two distinct components, a cap of takeoffs and landings and an allocation of the slot rights. Recall from our earlier discussion that assuming certain conditions, it was possible an efficient outcome could result regardless of the initial distribution of slot rights. With an active secondary market, it would be possible that trading of slot rights would lead to an efficient outcome.

While the FAA's imposition of hourly operational caps as a means of congestion management has faced little resistance, their decision to withdraw slots from current holders and auction them has been opposed by most legacy airlines, several

---

[54] A related concern is that implementing auctions and congestion pricing would reduce the incentive to increase capacity; however, the inefficient use of resources does not provide the justification for expansion. Furthermore, auctions would generate revenue that could be used for justified expansions.

[55] A concern that competitive pressures in the industry may be reduced stems from the possibility that such pricing would increase the entry barriers in the airline industry and thus protect legacy air carriers at the expense of smaller carriers and new entrants. Legacy air carriers also have access to resources to buy slots. On the other hand, the inefficiency of current operations should be weighed against these potential implications of, as well as the benefits from, market-based pricing.

[56] International flight times are somewhat inflexible, so the profitability of these flights would decline. There is a narrow time window both to leave New York for Europe and to leave Europe for New York to arrive on the same day because of European airports' slot rules and the North Atlantic Air Traffic Control system.

[57] Even if foreign carriers would be subject to congestion pricing or auctions, congestion pricing or auctions might have a large effect on domestic operations relative to international operations. Foreign carriers with primarily international operations at New York City area airports might endure the higher costs because international flights tend to generate more revenue per passenger than domestic flights. This would benefit foreign carriers relative to domestic carriers because they operate hubs in their home countries and would not be affected by the reduction of U.S. domestic feeder flights. Obviously, U.S. carriers would be adversely affected by the reduced number of such flights.

[58] See the comments of the ATA and Kasper and Lee (2009) before the FAA.

industry trade organizations, and various government entities.[59,60]

Not surprisingly, the FAA does not view the New York City market as generating an efficient allocation of slots. The FAA's concern is that the carriers serving LaGuardia have used some of their existing slots to deny their competitors access to the slots. To minimize the costs of this "babysitting" of the slots, the carriers use very small aircraft. If so, then there is unmet passenger demand at LaGuardia. This has prompted the FAA to encourage carriers to increase the average size of aircraft at the airport. Moreover, the ATA points to measures of concentration that indicate that LaGuardia is not dominated by a small number of carriers.[61]

In summary, the ATA believes that the FAA is attempting to solve a problem—a lack of competition that has caused consumer demand to be unmet—that does not exist. Thus, the auction and the related regulations are simply imposing costs on airlines operating at LaGuardia with no compensating gains.

On the contrary, the FAA views the totality of its proposal as necessary to generate an efficient and dynamically competitive environment not only at LaGuardia, but also at Kennedy and Newark. It views the slot auctions as primarily a long-run congestion management tool. Requiring established airlines to surrender a percentage of their slots would allow the FAA to increase competition inside the airports by providing new and limited carriers the ability to expand their operations. Without a forcible slot reallocation, the FAA argues, established airlines would have little incentive to ever voluntarily offer slots for sale or lease to potential competitors.[62,63]

CONCLUSION

Airport congestion continues as a major problem at many airports, including the three major airports in the New York City area. Economic theory suggests that this problem might be resolved via a quantity-based mechanism that includes a cap on the slots and a market-based allocation of the slots. However, determining the number of slots that yields an efficient level of congestion is a challenging task. In the present case, the FAA proposed an auction to sell a number of the slots. As with many public policy proposals, "the devil is in the details."

We have provided an extensive discussion of the proposal. Not surprisingly, many of the opponents believed they would be harmed by the proposed changes. Paying higher prices and/or incurring reductions in air transportation services are consequences that, not surprisingly, generate opposition from those likely to be harmed. Many of these opponents are able to avoid paying the full cost for their contributions to congestion. We have also conveyed the key features of the combinatorial auction that would be used to allocate slots. Without question, the FAA's plan to allocate some slots via an auction is more complex than the current system.

Although there is general agreement that a cap on the number of slots is appropriate, the use of the auction to allocate the slots has generated much controversy. The FAA viewed the auction as an essential feature of their plan, but auction opponents argued that it would have been unnecessary to ensure a competitive result. Moreover,

---

[59] One argument against the forcible slot reallocations is that the significant reduction in airport congestion intended by the final rules would result mainly from the hourly caps and not the application of auction proceeds realized from auctioning a marginal number of slots.

[60] While we cannot detail all objections here, they are well described in the FAA's final rule documents (*Federal Register*, 2008a,b).

[61] Based on one measure of concentration, the Herfindahl-Hirschman Index, LaGuardia ranked 34th of 40 airports with the largest number of domestic origin and destination passengers for the year ending June 2006. Even this ranking may tend to be overstated because of the consumer option to use other New York City area airports. See the comments of the ATA and Kasper and Lee (2009).

[62] While in the past airlines have been able to reallocate slots among themselves via the buy/sell rule, new and limited carriers have complained that, due to the lack of transparency requirements, established carriers could effectively shut them out of the market by arranging private transactions with other carriers, refusing to sell slots, or refusing to provide meaningful lease terms.

[63] Even though the overall number of slots available at each auction may be small relative to the total daily slot operations at each airport, given that the FAA intends to auction slots available during each peak operating hour, the awarding of even a small number of them will begin to establish the dollar value associated with the right to take off or land within a particular window of time. Moreover, once a monetary value is placed on a slot, it may be more difficult for established airlines to justify keeping those slots whose current best benefit is to keep competition at bay at the expense of clogging up the system.

and especially disconcerting to some airlines, the value capture feature of the proposed auctions would have required payments by airlines to the FAA for some slots that they previously received free of charge. Clearly, these opponents were not mollified by either the plan to use these proceeds for projects to mitigate delay and congestion or that auctions might be the least costly way of generating such funding. Without an auction, the FAA believed that the already-established carriers would have an unfair advantage. On the other hand, the opponents argue that these carriers are using the existing slots efficiently and that the existing secondary market works to ensure that slots are transferred to those airlines placing the highest values on the slots.

As of now, the opponents of using an auction to allocate slots have won the political battle. Whether a similar proposal will resurface and generate sufficient political support remains to be seen.[64]

REFERENCES

Aaronson, Scott. "The Limits of Quantum Computers." *Scientific American*, March 2008, pp. 62-69; www.cs.virginia.edu/~robins/The_Limits_of_Quantum_Computers.pdf.

Adar, Zvi and Griffin, James M. "Uncertainty and the Choice of Pollution Control Instruments." *Journal of Environmental Economics and Management*, October 1976, *3*(3), 178-88.

Bomkamp, Samantha. "Transportation Dept. Scraps NYC Slot Auctions." ABC News, May 13, 2009; http://abcnews.go.com/Business.wirestory?id=7576322.

Brueckner, Jan K. "Price vs. Quantity-Based Approaches to Airport Congestion." *Journal of Public Economics*, June 2009, *93*(5-6), pp. 681-90.

Caterinicchia, Dan. "DOT Proposes Slot Auctions to Reduce Congestion at NYC Airports." *USA Today*, May 16, 2008; www.usatoday.com/travel/flights/2008-05-16-new-york-slot-auctions_N.htm.

Coase, Ronald H. "The Problem of Social Cost." *Journal of Law and Economics*, October 1960, *3*, pp. 1-44.

Cohen, Jeffrey P. and Coughlin, Cletus C. "Congestion at Airports: The Economics of Airport Expansions." Federal Reserve Bank of St. Louis *Review*, May/June 2003, *85*(3), pp. 9-25; http://research.stlouisfed.org/publications/review/03/05/Cohen_Coughlin.pdf.

Cohen, Jeffrey P. and Coughlin, Cletus C. "An Introduction to the Two Rate Taxation of Land and Buildings." Federal Reserve Bank of St. Louis *Review*, May/June 2005, *87*(3), pp. 359-74; http://research.stlouisfed.org/publications/review/05/05/CohenCoughlin.pdf.

Cramton, Peter; Shoham, Yoav and Steinberg, Richard, eds. *Combinatorial Auctions*. Cambridge, MA: MIT Press, 2006.

Czerny, Achim I. "Congestion Pricing vs. Slot Constraints to Airport Networks." Working Paper No. 2006-02, Berlin University of Technology, Center for Network Industries and Infrastructure, 2006.

Dubner, Stephen J. "Want to Fix New York Air Congestion? Shut Down LaGuardia." *The New York Times*, May 14, 2009; http://freakonomics.blogs.nytimes.com/2009/05/14/want-to-fix-new-york-air-congestion-shut-down-laguardia/.

Economist. "Cap and Trade, with Handouts and Loopholes." May 23, 2009, pp. 33-34.

Federal Aviation Administration. "Auction Procedures for Allocating Slot at LaGuardia, John F. Kennedy International, and Newark Liberty International Airports." September 15, 2008a; www.fbo.gov/index?s=opportunity&mode=form&id=0fd39be76893145ad96b6c0b959203d1&tab=core&_cview=1&cck=1&au=&ck=.

Federal Aviation Administration. "2009 New York Slot Auctions Bidders Seminar" (slide presentation), December 5, 2008b; www.faa.gov/about/office_org/headquarters_offices/aep/ny_auctions/media/bidder-seminar-2009-new-york-slot-auctions-2.pdf.

[64] See Levine (2009) for a proposal that attempts to generate an efficient allocation of slots using an auction, while also attempting to overcome the objections to the FAA's proposal.

Federal Register. "Congestion Management Rule for John F. Kennedy International Airport and Newark Liberty International Airport; Final Rule." October 10, 2008a, *73*(198), pp. 60544-60571; www.fly.faa.gov/ecvrs/circulars/FR_JFK_and_EWR_final_rule_101008.pdf.

Federal Register. "Congestion Management Rule for LaGuardia Airport; Final Rule." October 10, 2008b, *73*(198), pp. 60574-60601; www.fly.faa.gov/ecvrs/circulars/FR_LGA_Final_Rule_101008.pdf.

Federal Trade Commission. "Comment of the Staff of the Bureau of Economics of the Federal Trade Commission—In the Matter of Congestion Management Rule for LaGuardia Airport." Docket No. FAA-2006-25709, Notice No. 08-04; www.ftc.gov/os/2008/06/V080015comment.pdf.

Holland, Jesse J. "Court Blocks FAA Auction of Airport Slots," ABC News, 18 December 2008; http://abcnews.go.com/print?id=6422027.

Kasper, Daniel and Lee, Darin N. "Supplemental Notice of Proposed Rulemaking Congestion Management Rule for LaGuardia Airport." Comments before the Federal Aviation Administration, Docket No. FAA-2006-25709, June 16, 2009; http://lecgaviation.com/thought_leadership/LECG%20Comments,%20June%2016.pdf.

Levine, Michael E. "Airport Congestion: When Theory Meets Reality." *Yale Journal on Regulation*. Winter 2009, *26*(1), pp. 37-88.

Morrison, Steven A. and Winston, Clifford. "The Effect of FAA Expenditures on Air Travel Delays." *Journal of Urban Economics*, March 2008, *63*(2), pp. 669-78.

New York Aviation Rulemaking Committee Report. December 13, 2007; www.faa.gov/library/reports/media/NY%20ARC%20Final%20Report.pdf.

Stavins, Robert N. "Correlated Uncertainty and Policy Instrument Choice." *Journal of Environmental Economics and Management*, March 1996, *30*(2), pp. 218-32.

Telser, Lester G. "The Usefulness of Core Theory in Economics." *Journal of Economic Perspectives*, Spring 1994, *8*(2), pp. 151-64.

Tietenberg, Tom. "Tradable Permit Approaches to Pollution Control: Faustian Bargain or Paradise Regained?" in M.D. Kaplowitz, ed., *Property Rights, Economics, and the Environment*. Stamford, CT: JAI Press, 2000, pp. 155-76.

Weitzman, Martin. "Prices vs. Quantities." *Review of Economic Studies*, October 1974, *41*(4), pp. 477-91.

Whalen, W. Tom; Carlton, Dennis W.; Heyer, Ken and Richard, Oliver. "Proposal for a Market-Based solution to Airport Delays." Economic Analysis Group of the Antitrust Division of the U.S. Department of Justice Discussion Paper, EAG 07-14, October 2007; www.usdoj.gov/atr/public/eag/227380.pdf.

The Commercial Paper Market, the Fed, and the 2007-2009 Financial Crisis

Richard G. Anderson and Charles S. Gascon

Since its inception in the early nineteenth century, the U.S. commercial paper market has grown to become a key source of short-term funding for major businesses, with issuance averaging over $100 billion per day. In the fall of 2008, the commercial paper market achieved national prominence when increasing market stress caused some to fear that, given its size and importance, the market's failure would sharply worsen the recession. The Department of the Treasury and Federal Reserve enacted programs targeted at providing credit and liquidity to restore investor confidence. The authors review the history of the commercial paper market, describe its structure and key relationships to money market mutual funds, and present a detailed discussion of the crisis in the market, including the resulting Federal Reserve programs. (JEL G01, G24, E52)

Federal Reserve Bank of St. Louis *Review*, November/December 2009, *91*(6), pp. 589-612.

For both small and large American businesses, commercial paper (CP) issuance is an important—and often lower-cost—alternative to bank loans as a means of short-term financing. Paper is sold in different forms: Some paper is sold unsecured (that is, without specific collateral), while other paper is secured by bank-issued letters of credit or pools of assets, including a firm's receivables. The funds raised through CP issuance have a variety of uses, including payroll and inventory finance.

CP is important for investors as well. Larger investors, including institutions, directly purchase CP as a short-term, low-risk investment. For smaller investors, money market mutual funds (MMMFs) intermediate between larger-denomination CP and the liquid, smaller-denomination shares that they issue to the public.[1]

Since the inception of the CP market in the early nineteenth century, it has grown such that today CP issuance exceeds that of Treasury bills. The early years of the CP market were dominated by issuers in the nonfinancial sectors of the economy, including transportation and utility companies, who borrowed by issuing CP to wealthy individuals, other businesses, and financial institutions. By the twentieth century, as the demand for durable goods rose and consumers began purchasing items on credit, the CP market became dominated by financial issuers. The rise of MMMFs during the 1970s boosted the growth of CP by (indirectly) allowing small investors access to CP investments. During the 1980s, the CP market began to develop into its current form, particularly with the creation of the asset-backed commercial paper (ABCP) conduit.

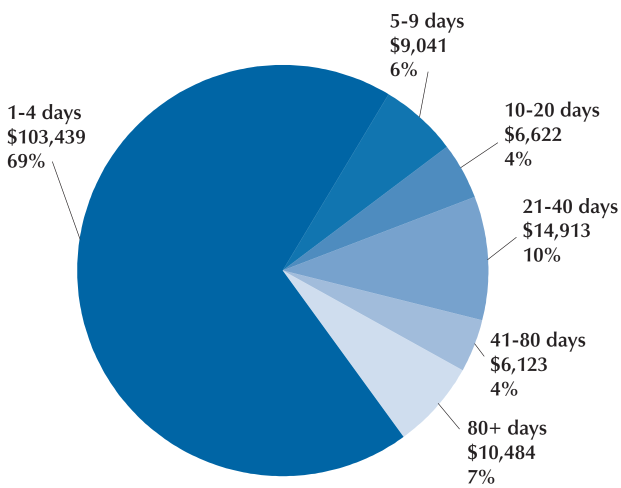
During the autumn of 2008, some feared that stress on—and potential failure of—the CP market

[1] MMMFs purchase low-risk, large-denomination securities such as commercial paper and government securities, and issue "shares" at $1 per share.

Richard G. Anderson is a vice president and economist at the Federal Reserve Bank of St. Louis and a visiting scholar at the School of Business, Aston University, Birmingham, United Kingdom. Charles S. Gascon is a senior research associate at the Federal Reserve Bank of St. Louis.
The authors thank Standard & Poor's Financial Services LLC, Moody's Investors Service, and Fitch Ratings for their assistance in the preparation of this article and permission to cite their reports. Special thanks is extended to Everett Rutan for his comments on earlier versions.

Figure 1

Commercial Paper Issuance by Maturity, 2008 (average, $ billions)



SOURCE: Federal Reserve Board, Volume Statistics for Commercial Paper Issuance.

might sharply worsen the recession. In response, the Treasury and Federal Reserve enacted programs to restore stability. These programs focused on enhancing market liquidity, not on removing default risk from the market.

THE STRUCTURE OF THE COMMERCIAL PAPER MARKET

In its traditional form, CP is an unsecured promissory note issued by a business (either financial or nonfinancial) for a specific dollar amount and with maturity on a specific date.[2] Companies issue CP as a low-cost alternative to bank loans, as it is exempt from Securities and Exchange Commission (SEC) registration.[3] CP is generally issued in large denominations of ($100,000 or more). The maturity on CP averages 30 days but may range up to 270 days. Proceeds from CP issuance are used to finance "current transactions," including meeting payroll obligations, and funding current assets, such as managing receivables or inventories. Figure 1 displays the maturity composition of CP issued in 2008; that year, the majority of CP had a maturity of less than 1 week.

Similar to Treasury bills, CP is typically issued at a discount, meaning that the buyer pays less than face value and receives face value at maturity: The "interest" is equal to the face value minus the purchase price. Although CP is issued at short maturities to minimize interest expense, many issuers *roll over* CP by selling new paper to pay off maturing paper. Because of modest credit risk, yields on CP are slightly higher than on Treasury bills of similar maturity. Large denominations and short maturities typically limit the CP market to large institutional investors, such as MMMFs.

CP generally is classified in three broad (but overlapping) categories: nonfinancial, financial, and asset-backed. Further, CP may be classified as being sold with the assistance of a CP dealer (dealer placed) or without (directly placed). Traditional nonfinancial and financial paper, respectively, are unsecured short-term debt issued by highly rated corporations, including industrial firms, public utilities, bank holding companies, and consumer finance corporations. ABCP, on the other hand, is more complicated. The simplest description of ABCP is as a form of securitization: As the name implies, it is CP with specific assets attached. In financial industry jargon, ABCP is issued by "conduits." Conduits are structured to be bankruptcy remote and limited in purpose.[4] Each conduit includes a special-purpose vehicle (SPV) that is the legal entity at the center of the program and a financial adviser (usually a commercial or investment bank) that manages the program and determines the assets to be purchased

---

[2] The exception being ABCP, which is CP with specific assets attached as collateral. We discuss this type of CP in more detail below.

[3] CP is exempt from SEC registration if the following three criteria are met: (i) the maturity of the paper is less than 270 days, (ii) notes must be of a type not ordinarily purchased by the general public, and (iii) issues must be used to finance "current transactions" (Hahn, 1998).

[4] *Bankruptcy remote* refers to all participants in a conduit agreeing not to force the SPV into bankruptcy prior to a year and a day after issuance of the conduit's most recent CP; this ensures redemption of all CP. *Limited purpose* refers to issuance of CP as its sole business.

Table 1

Asset Composition of Multiseller Conduits (approx.)

	2002	2006	2007	2008
Consumer Assets				
Credit Card Receivables	41.9	44.2	44.4	45.9
Auto Loans and Leases	12.3	13.8	16.0	17.8
Credit Cards	14.9	10.5	13.1	12.7
Student Loans	1.0	4.2	7.2	8.5
Residential Mortgages	8.2	11.3	4.8	4.4
Other Consumer	5.5	4.3	3.1	2.4
Commercial Assets	51.5	43.3	54.8	44.0
Trade Receivables	14.6	15.4	13.9	14.0
Commercial Loans and Leases	6.6	12.4	13.6	12.9
Equipment Loans and Leases	10.6	3.4	3.6	4.1
Other Commercial	19.6	12.1	14.7	13.0
Securities	6.6	12.5	9.7	10.1
Total (percent)	100	100	100	100

SOURCE: Moody's ABCP Query: Data are the share of total outstanding as of year-end.

and the ABCP paper to be issued. The owner of the conduit receives nominal dividend payments; and because the SPV does not generally have any employees, fees are paid to an administrator (normally a bank) to manage the flow of CP and funds.

To investors, ABCP programs are less transparent than traditional unsecured corporate CP—the SPV is an opaque entity that holds assets that are unknown to the purchaser of the ABCP. Performance of the ABCP depends on the skill of the bank adviser, which essentially is saying to ABCP purchasers: "Trust us with your funds and we will invest them for you." The degree of disclosure in the market varies widely. Some multiseller[5] conduits provide investors with at least a list of assets, liquidity enhancements, and performance history, whereas other more-complex conduits provide very limited disclosure. During normal times, yields on ABCP have been approximately 75 basis points greater than yields on traditional unsecured CP.[6] This spread is a continued mystery with numerous explanations: Why should CP with assets attached as collateral pay a higher yield than CP with no such collateral? Some have suggested this spread exists because ABCP is (indirectly) being issued by firms unable to directly issue their own CP. However, this does not explain why ABCP issued by a conduit sponsored (and insured) by a bank would have a higher yield than CP directly issued by the same bank lacking any collateral. Moody's (2009) attributes the yield premium specifically to the lack of transparency, noting that traditional CP is relatively easily understood, while ABCP is issued by an unfamiliar SPV with assets from anonymous sellers.

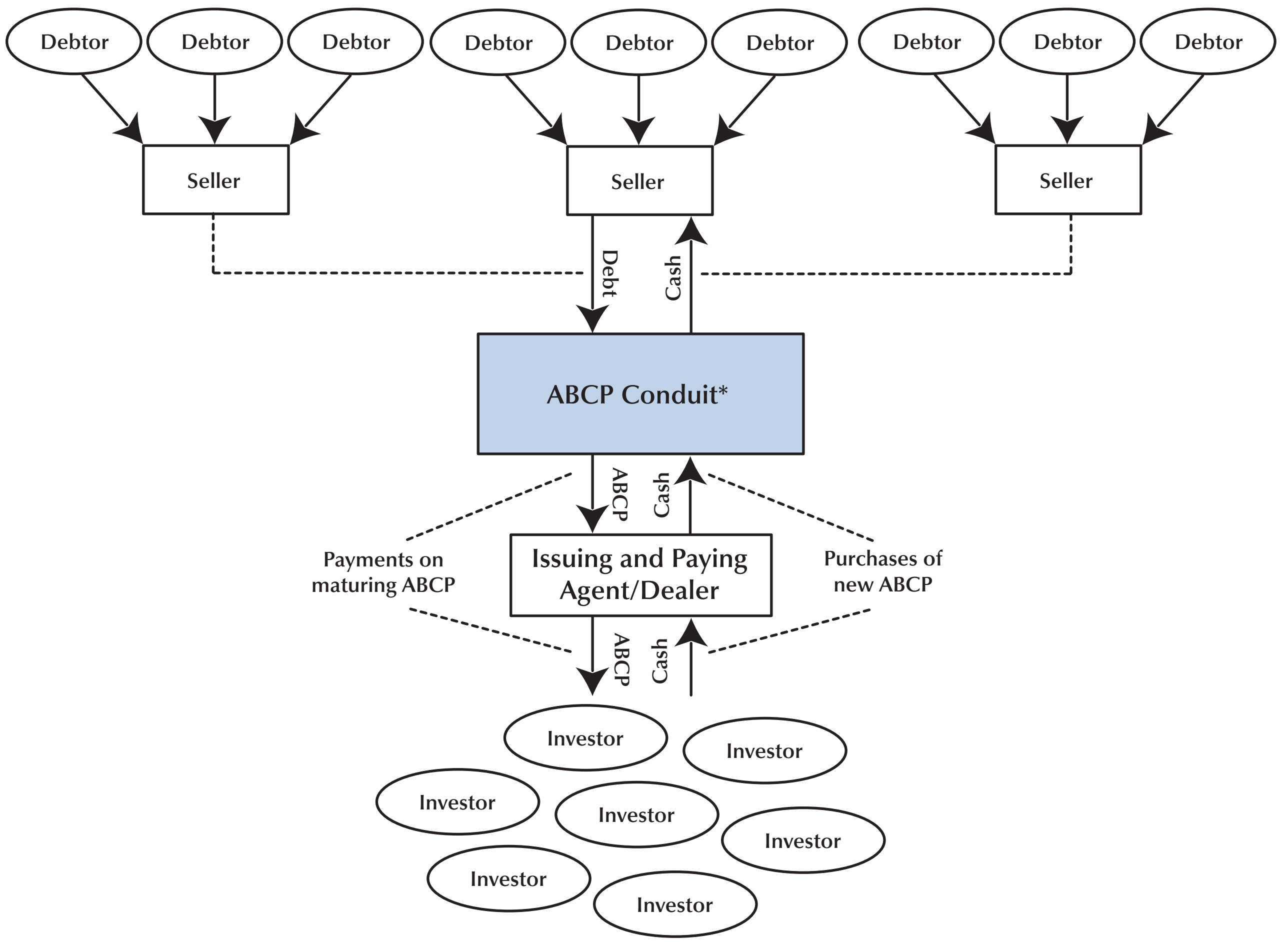
Because of its role in the 2008 credit crisis, some additional discussion of ABCP is valuable. CP conduits, in their structure, are classic financial intermediaries: They purchase one or more types of financial assets and issue ABCP in their

5 A bank-sponsored multiseller conduit backs its CP with a diverse pool of assets, as opposed to a single-seller conduit (e.g., a conduit sponsored by General Motors Acceptance Corporation), which backs its CP with pools of specific assets (e.g., auto loans).

6 This spread can be calculated using rates on AA-rated CP reported in the Federal Reserve Board volume statistics on CP issuance. The spread changes depending on the issue type (financial or nonfinancial) and maturity chosen.

Figure 2

ABCP Conduits—Banking Without a Bank



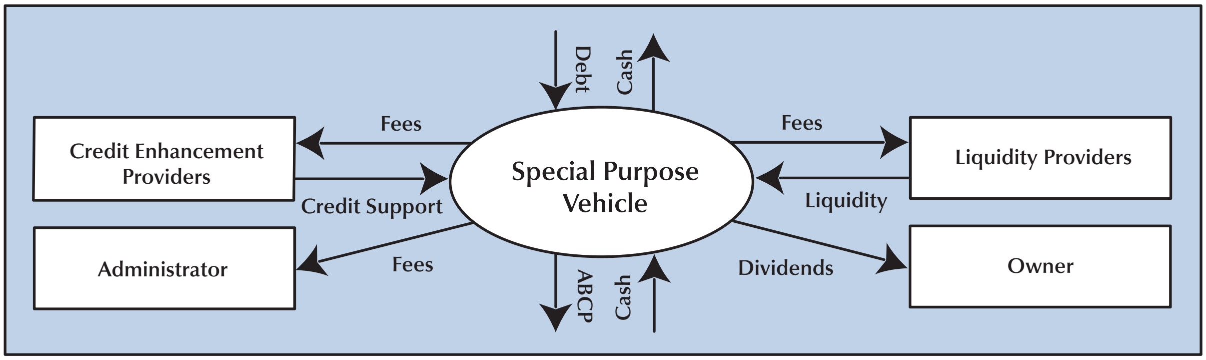
NOTE: *See Figure 3 for additional details.

own name. Their business purpose is to arbitrage risk and rate spreads between the assets they purchase and the liabilities they issue. ABCP conduits come in many types. The largest type is multiseller conduits. The major cost to a multiseller ABCP conduit is the insurance of risk. To insure risk, conduits pay fees to a liquidity provider (rollover risk) and a credit-enhancement provider (default risk).

ABCP issued by multiseller conduits is often used to finance the purchase of consumer and commercial assets (Table 1). A firm that sells receivables to a conduit frees its own funds and reduces its need to borrow. Large, creditworthy firms typically sell their paper directly to investors via an agent or dealer, a practice referred to as a "single-seller conduit." Smaller firms, however, find this method costly, and prefer to operate via a multiseller ABCP conduit in which the firm sells its debts to a bank-advised SPV which, in turn, sells ABCP to investors. Large firms may also use multiseller ABCP conduits as an additional source of liquidity in cases when they have a quick turnaround on trade receivables that are not large enough to warrant "traditional" CP issuance. Unrated firms may also lower borrowing costs

Figure 3

Typical ABCP Conduit



SOURCE: Adapted from Moody's (2003).

via multiseller conduits by paying a spread to the conduit sponsor.

Figures 2 and 3 outline the securitization of assets through a multiseller ABCP conduit. At the top of the diagram are debtors (i.e., individuals or businesses) who borrow money from sellers (i.e., mortgage lenders or banks). In the center of Figure 2 is the ABCP conduit; its SPV is illustrated in Figure 3. The ABCP SPV purchases, from the sellers, the debt at some price less than face value. This overcollateralization (or "haircut") provides an equity cushion to CP investors. Because finding suitable investors may be costly, the ABCP SPV has a relationship with a dealer (i.e., an investment bank), who suggests a price and finds suitable investors.

Because the maturity on the CP is shorter than the maturity on the original loans, the ABCP conduit will roll over the maturing CP to pay investors. As with any CP program, rating agencies require ABCP conduits to obtain liquidity backstops on each transaction to assume much of the rollover risk. Liquidity providers will normally provide funds on non-defaulted assets. Because there is always a risk that some debtors fail to make payments, investors require additional program-wide credit enhancements, generally in the form of a bank letter of credit or insurance company surety bond on some fraction of the maximum program size.[7] Normally, these agreements require payment by the provider once other sources of funds have been exhausted. In many cases the administrator is the same bank providing the liquidity and credit enhancements. As a result, the credit rating of the conduit is closely related to the credit rating of liquidity and credit providers, as well as the reputation of the managing party.

Table 2 shows the composition of CP outstanding by issuer and placement type. The top two rows report the average amount outstanding in 2001 and 2008, and the bottom two rows report shares as a percentage of the total. For each issuer, the average amount outstanding is disaggregated by the placement type. For example, during 2001, financial CP outstanding averaged $617.0 billion (41.4 percent of the total), of which $336.5 billion was placed by dealers (22.6 percent of the total) and $280.6 billion was directly placed (18.8 percent of total). On average, 90 percent of outstanding CP was either ABCP or financial CP in 2008.

A firm ordinarily requires a dealer to place its paper if it lacks the name recognition necessary

[7] This really applies only to bank-sponsored multiseller programs. Securities arbitrage programs have recently added letters of credit, while structured investment vehicles use overcollateralization or sell subordinate notes. Single-seller conduits tend to use overcollateralization.

Table 2

Commercial Paper Outstanding by Issuer and Placement

	Financial			Nonfinancial			Asset-backed			All types
	Dealer	Directly	Total	Dealer	Directly	Total	Dealer	Directly	Total	Total
Total (average), $ billions										
2001	336.5	280.6	617.0	205.9	38.5	244.4	500.8	127.6	628.4	1489.8
2008	552.2	231.5	783.7	174.6	17.1	191.7	663.1	100.4	763.6	1739.3
Share (percent)										
2006	22.6	18.8	41.4	13.8	2.6	16.4	33.6	8.6	42.2	100.0
2008	31.7	13.3	45.1	10.0	1.0	11.0	38.1	5.8	43.9	100.0

SOURCE: Federal Reserve Board, Volume Statistics for Commercial Paper Issuance.

to attract investors or if its funding requirements either are too limited or infrequent to warrant building its own distribution system. Direct issuers of CP, most of them traditional issuers, borrow in sufficient size and frequency that the costs of developing an in-house distribution system are less than the costs of placing paper through a dealer. For nonbanks, an in-house system may become profitable if CP issuance reaches $500 million or more. Mostly the major finance companies and large banking organizations that also distribute wholesale liabilities (such as certificates of deposit [CDs]) place their paper directly. Only a few nonfinancial firms are direct issuers of paper.

Many companies build close relationships with their dealers: If a company is willing to sell its paper at the dealer's suggested price, the dealer will agree to purchase unsold paper. Yet, relationships with dealers may be problematic and are not *explicitly* guaranteed. Market intelligence suggests that dealer relationships for multiseller conduits were "strained to the breaking point" during the fall of 2007, and "collapsed" for many single-seller and securities arbitrage conduits. Generally, according to dealer reports in Stigum and Crezcenzi (2007), "competition among dealers is fierce...at a 70 percent utilization rate you maybe break even or are losing a bit of money...[at a] 90 percent utilization rate you begin to make real profits." Dealers charge clients a fee that is less than one-eighth of 1 percentage point, which in 2008, translated into roughly $150 million in daily fees on $120 billion of CP issued daily.[8]

THE DEVELOPMENT OF THE COMMERCIAL PAPER MARKET

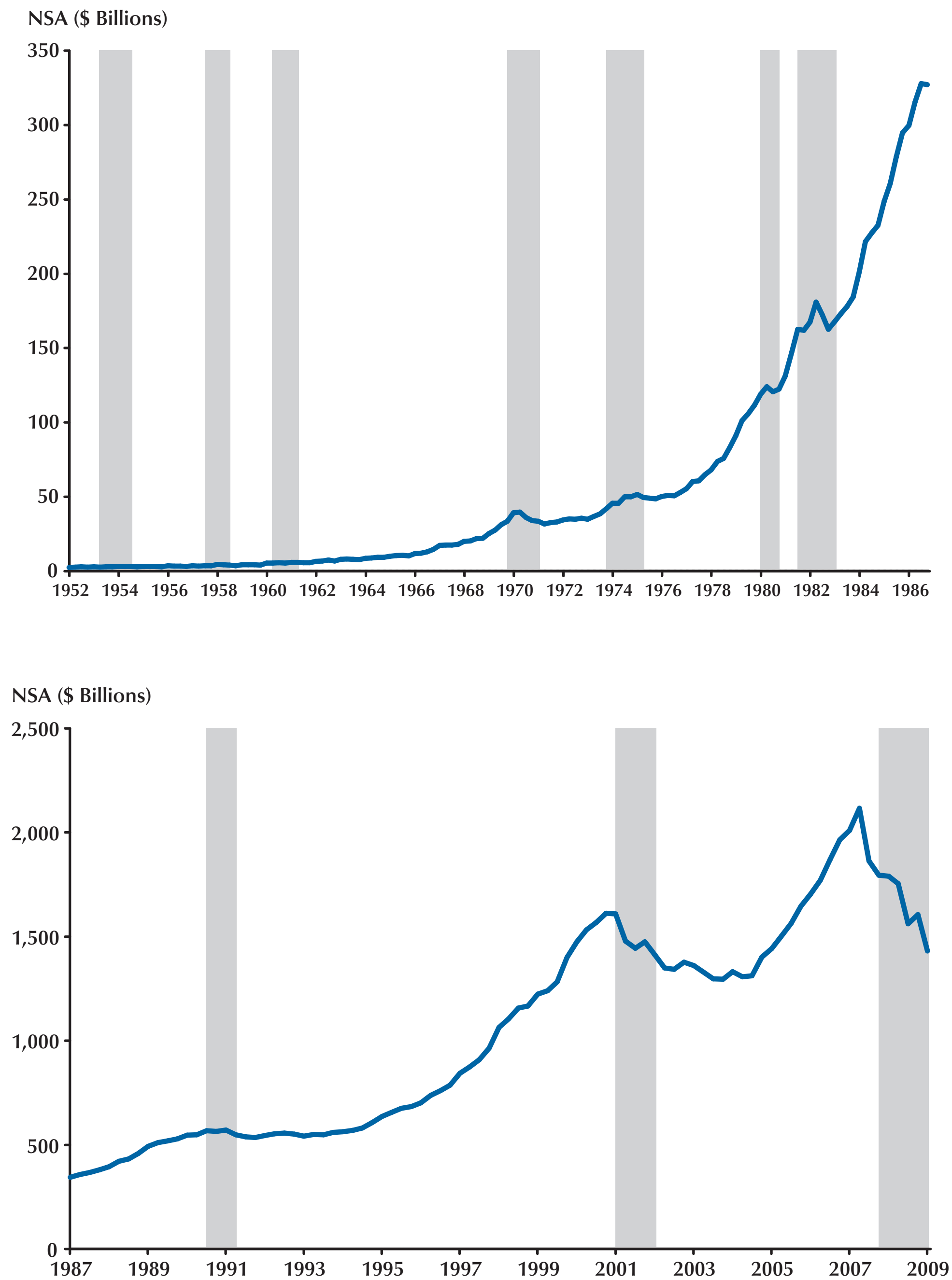
In the early years of the CP market, the nineteenth century, nonfinancial firms (including textile mills and railroad companies) were the major issuers of paper. By the early twentieth century, particularly following the founding of General Motors Acceptance Corporation (GMAC) in 1919, the CP market expanded to include financial paper. After World War II, increased sales of durable goods on credit (especially televisions and automobiles) encouraged expansion of consumer finance companies and, in turn, the CP market (Stigum and Crezcenzi, 2007, Chap 3). Later, even for business purposes such as financing inventory and raising cash for current operating expenses, CP increasingly replaced bankers' acceptances as the instrument of choice for short-term financing.

Figure 4 shows the trend in the amount of CP outstanding (all issue types) since 1952. Prior to 2000, the CP market grew steadily as both borrow-

[8] Data on daily issuance of dealer paper are not available; the calculation assumes 80 percent (the percentage of total outstanding issued by dealers) of the average daily issuance is placed by dealers. The dealer fee of 0.0125 percent is from Stigum and Crezcenzi (2007, p. 989).

Figure 4

Quarterly Commercial Paper Outstanding (1952-1986 and 1987-2009)



SOURCE: Federal Reserve Board, Flow of Funds, Table L.208.

Table 3

Major Holders of Commercial Paper*

	Percent				$ Billions
	1952-1971	1972-1991	1992-present	2008	2008
Money Market Mutual Funds	—	18.2	34.7	38.5	615.6
Funding Corporations	3.3	11.5	12.9	23.0	367.5
Foreign Sector	15.6	4.6	9.5	14.6	233.2
State & Local Governments	—	—	8.1	7.7	123.8
Security Brokers & Dealers	—	4.7	3.0	4.1	65.7
Mutual Funds	3.3	2.6	5.7	3.3	52.0
Life Insurance Companies	2.2	7.5	5.2	2.7	42.8
Private Pension Funds	—	8.2	2.8	2.3	36.9
State & Local Gov. Retirement Funds	—	0.8	3.2	2.0	31.9
Commercial Banking	18.4	7.5	0.4	0.9	15.0
Nonprofit Organizations	42.7	23.3	8.2	0.7	10.4
Nonfarm Nonfinancial Corporate Business	14.0	9.0	4.0	0.3	4.7
Monetary Authority	0.5	—	—	—	0.0
Savings Institutions	—	1.2	—	—	—
Credit Unions	—	0.1	0.1	—	—
GSEs	—	0.6	2.1	—	—
Total (percent)	100.0	99.7	100.0	100.0	
Total ($ billions)	13.3	262.9	1,234.4	1,599.5	1,599.5

NOTE: *Data reported are for *open market paper*, which contains both CP and bankers acceptances. CP comprises 85 percent of open market paper over the sample and 99 percent since 1998.

SOURCE: Federal Reserve Board, Flow of Funds, Table L.208.

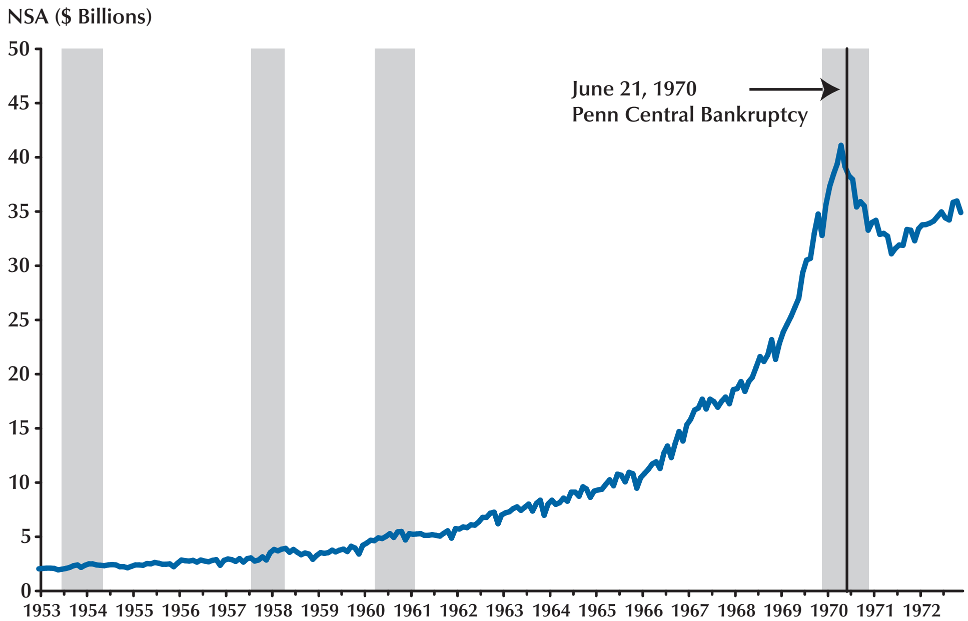
ers and investors shifted into CP from alternative money market instruments, including Treasury bills, bankers' acceptances, and CDs.[9] The downturn that took place in the 2000 is discussed further below. In 1970, CP comprised only one-quarter of the dollar volume of outstanding money market instruments; in 2006, it comprised two-thirds (Stigum and Crezcenzi, 2007, p. 967).

The introduction of MMMFs in 1971 had a large and long-lasting impact on the CP market. Fueled initially by rising demand for consumer durables, growth of the CP market was ignited in the 1970s by widespread investor enthusiasm for MMMFs. For savers and investors, MMMF shares were an attractive alternative to bank deposits; for corporate borrowers, CP was an attractive alternative to bank loans. Assets of MMMFs increased sixfold between 1980 and the end of 1991 (Table 3). During the period spanning 1972-92, MMMFs on average held 18.2 percent of all outstanding CP; in 2008, it was almost 40 percent. The increase in MMMF holdings was not steady: During three years (1978-81), the share of CP held by MMMFs soared to 32 percent from less than 1 percent, subsequently remaining near 30 percent. At year-end 1991, the MMMF industry held about one-third of all CP outstanding and was the largest single investor. Holdings of CP by foreign investors, on the other hand, have increased gradually since the early 1990s. (In Table 3, the share held by funding corporations

[9] See Anderson (2009a,b) for a discussion of bankers' acceptances. Today, the bankers' acceptance market is moribund, overtaken by CP.

Figure 5

Monthly Commercial Paper Outstanding (1953-1974)



SOURCE: Board of Governors (1976) and Federal Reserve Board, Commercial Paper Statistics.

is the residual, holdings not accounted for by the other categories.)[10]

Problems in the Early Years: The Penn Central Collapse

Penn Central railroad was a major issuer of CP, with approximately $84 million outstanding in the summer of 1970. As the company's cash flows dwindled, debt holders pushed for government assistance that would have allowed Penn Central to repay maturing CP. The assistance plan failed, and on June 21, 1970, Penn Central filed for bankruptcy (Calomiris, 1994). The bankruptcy of Penn Central rattled the CP market. The economy was already in recession, and the financial health of the company had *apparently* deteriorated in a matter of months. Market participants became worried that other highly rated CP issuers could be in a similar position.

In the years before the collapse, the CP market had experienced rapid growth and appeared to be isolated from economic downturns. Figure 5 indicates that total CP outstanding did not decline during either of the prior recessions. Because the market had not previously experienced such stress, lenders were uncertain of potential spillover effects of the bankruptcy, including the inability to roll over existing paper at maturity. The unwillingness of the Congress and the Federal Reserve to ensure payment of Penn Central's debt left creditors facing substantial losses. In response to the crisis, the Fed encouraged member banks to borrow at the discount window and make loans to CP issuers.[11] According to reports at the time,

[10] Funding corporations consist of four types of financial institutions and entities: (i) subsidiaries of foreign banks that raise funds in U.S. markets and transfer proceeds to foreign banking offices in the United States; (ii) subsidiaries of foreign banks and nonbank financial firms that raise funds in the United States and transfer them to a parent company abroad; (iii) nonbank financial holding companies; and (iv) custodial accounts for reinvested collateral associated with securities-lending operations.

[11] See Calmoris (1994) for additional details.

discount window borrowing to finance CP rollovers reached $500 million in the weeks following the collapse. The actions of the Fed assured financial markets that the liquidity needed to meet obligations would be available. Nevertheless, outstanding CP declined by 21 percent during subsequent quarters before turning upward in the second quarter of 1972. Eventually, the market regained its confidence—CP grew steadily for the next three decades. However, after the crisis CP issuers were more reserved and began securing lines of credit in case of market disruptions.

The Adolescence of the U.S. Commercial Paper Market: 1980-1992[12]

The U.S. CP market matured during the 1980s. At the beginning of the decade, issuance was primarily by a small number of large, prominent, and creditworthy companies. During the decade, everything changed. The market's size grew fivefold. New issuers and dealers arrived, while some older issuers disappeared. New forms of paper were introduced, most importantly ABCP.

In the 1980s, corporate borrowers faced paying relatively high rates (compared with historical experience) on both long-term funds (bonds) and bank loans, owing in part to robust economic growth. A less costly alternative was issuance of CP, which grew rapidly. Many new issuers were attracted to the market, including smaller U.S. corporations, foreign corporations, and foreign financial institutions. The development of a market in currency swaps allowed foreign borrowers to combine U.S. dollar-denominated CP issuance with swaps so as to create liabilities in other currencies. ABCP also came into general use, providing off-balance-sheet financing for trade and credit card receivables. Finally, the growth of MMMFs, coupled with a shift in the composition of their investments toward CP, made them the largest single source of funds to the market (see Table 3).

A series of defaults on CP that began in 1989 caused tighter regulations to be imposed on MMMF holdings of medium-grade paper. Heightened investor concerns effectively forced many medium-quality issuers to cut back sharply on their use of the CP market. Increasing costs also changed the role of banks in the CP market. Financial stress at banks became manifest in the pressure from markets and regulators to increase their capital levels, which in turn increased their costs of providing letters of credit and backup liquidity to the CP market. Partially offsetting this effect, in terms of overall CP market volume, were efforts of banks to increase loan rates and margins on loans. Growth of the CP market was neither smooth nor painless. The composition of firms issuing CP changed as defaults reduced investor appetite for medium-grade paper. At times, issuers returned to banks, finding bank loans less expensive than CP.

Defaults of CP are rare. Between 1971 and mid-1989, no defaults occurred in U.S. CP except for the litigation-driven default by Manville Corporation in 1982 (Post, 1992, p. 888). In mid-1989, the U.S. CP market was hit with three defaults; four more followed in 1990. Because fund advisers injected capital to cover the shortfalls, investors incurred no losses. The SEC subsequently tightened Rule 2a-7 to generally require two ratings on CP held by money funds and to limit a fund's holdings of a single firm's paper (p. 889). Growth of the paper market slowed thereafter, and some medium-grade issuers found borrowing at banks less expensive.[13]

> The financial markets calmed after 1990 and were capable of handling the funding needs of medium-grade firms. Medium-grade issuers successfully tapped bank lines of credit or their CP dealers, while ABCP absorbed some of the needs of these firms and grew rapidly. But investors remained wary of medium-grade paper. Interest rates on it spiked again both at midyear and at year-end 1991 because many investors did not want to show such holdings on their published financial statements. The June 1991 default of Columbia Gas, a second-tier issuer, renewed concerns about the safety of medium-grade paper (p. 889).

[12] This section is based on Post (1992).

[13] New Basel Accord risk-based capital guidelines for banks, adopted in 1988, would become effective at year-end, and market participants grew increasingly uncertain about the capacity of banks to honor all their loan commitments. As a result, rates paid on CP, even by highly rated firms, jumped in December 1990. This proved, however, to be the point of maximum stress.

Issuers at the end of the 1980s differed greatly from those at the beginning. At the end of 1989, about 1,250 corporations and other entities had paper programs in the U.S. CP market, 500 more than in 1980. Many new issuers were foreign firms and smaller, lesser-known U.S. firms; previously, CP issuers almost uniformly were large, well-known U.S. corporations.

The Growth in Dealer-Placed Financial Paper. Throughout the life of the CP market, methods of issuance have continued to evolve. During the 1980s, direct issuers expanded rapidly. Early in the decade, approximately 60 percent of all CP was sold directly by issuers to investors (p. 883). Among the more important issuers were large finance companies; these grew rapidly after the Economic Recovery Tax Act of 1981 promoted business use of leasing.

> Bank holding companies continued to use the CP market to support parent company operations, including leasing and lending by non-bank subsidiaries. By the end of the decade, outstanding paper placed directly by financial firms surpassed $200 billion, more than triple the level at the start of the decade…[Yet, even faster growth was experienced by firms that used dealers for distribution.]…By 1989, dealer-placed paper accounted for 60 percent of all CP outstanding, up sharply from about 40 percent at the start of the decade (p. 883).

In part, the growth was supported by Federal Reserve Board rulings in 1986 and 1987 that authorized certain so-called Section 20 subsidiaries of bank holding companies to deal in CP to a limited extent; by year-end 1991, these subsidiaries accounted for about 14 percent of outstanding dealer-placed paper. And by December 1990, dealer-placed financial CP outstanding surpassed the amount of directly placed financial CP (p. 884).

The increased share of dealer-placed paper also reflected, in part, the changing composition of issuers: Dealers were required for the aggressive marketing needed to package and sell new issuers and new types of CP programs.

> During the mid- to late 1980s, the presence of foreign financial institutions in the U.S. market grew, and these firms generally required dealer assistance to promote their names to U.S. investors. By year-end 1991, these firms had outstanding CP in excess of $110 billion, slightly more than half of all dealer-placed financial paper. Highly rated foreign banks (or their U.S. subsidiaries) accounted for 55 percent of this paper (pp. 884-85).

The Growth in Guaranteed Paper. The growth in guaranteed paper is described by Post (p. 884).

> The share of CP programs that were fully (100 percent) enhanced by credit guarantees—often bank letters of credit—from highly rated third parties grew dramatically in the first half of the decade. In fact, programs with such credit enhancements accounted for about all the net increase in the number of CP issuers rated by Moody's over that period. Presumably, most of these programs were small because their outstanding CP accounted for less than 10 percent of all outstanding paper.
>
> Because investors in such paper rely on the guarantor, rather than the issuer, to make payment in full upon maturity of the paper, the paper carries the rating of the guarantor. Whereas traditional issuers entered the market on the strength of their own credit quality (or that of their parent), many of the new CP programs of the first half of the 1980s gained access to the market on the strength of guarantees by unrelated entities.

Introduction of the Asset-Backed Commercial Paper Conduit. The decade's second innovation, and perhaps its most important, was the introduction of the bank-advised ABCP conduit in 1983.[14] The structure of the typical multiseller ABCP conduits was discussed previously. When considered solely by their economic functions, such conduits, essentially, are regarded as "banks without banking charters." The motives at the time of their introduction are well described by Post (1992, p. 886):

> The development of the asset-backed sector of the CP market arose from several factors. U.S. banking organizations saw an opportunity to generate fee income from potential participants in their programs—many of

[14] See Kavanaugh, Boemio, and Edwards (1992).

> which were the same investment-grade firms that they had lost as loan customers to the CP market. These banking organizations also became more familiar with asset securitization. This similarity resulted, in part, from increased market and regulatory pressure to increase their capital ratios. Asset securitization, and asset-backed CP in particular, permitted banks to channel would-be borrowers to funding off of bank balance sheets. Another factor was that financial markets became increasingly familiar with, and thus more willing to accept, programs that required structuring, such as those with credit guarantees. Dealers saw opportunities to market asset-backed programs to companies seeking to increase liquidity or to reduce leverage, regardless of size or rating. Moreover, they already had proved successful in marketing lower-rated firms to the CP market via guaranteed programs and realized that a pool of potential business existed in companies that were too small to tap the CP market through their own guaranteed programs. Thus, banking organizations formed bank-advised asset-backed programs, relying on dealers.

ABCP conduits increased in number from 3 in 1985 to 89 by year-end 1991. Between 1990 and 1991, ABCP programs accounted for virtually all the increase in domestic CP issuers. By year-end 1991, ABCP accounted for about 9 percent of all outstanding CP.

In circumstances reminiscent of the 2008 credit crisis, credit problems at sponsoring bank holding companies slowed the growth of ABCP paper during the 1989-92 credit crisis. Cantor and Roriques (1994) report that the perceived credit risk of CP increased as the number of defaults on CP "soared" (p. 171) and the number or downgrades outpaced the number of upgrades between 1988 and 1989 (p. 194). ABCP conduit ratings were downgraded as large loan losses, and the need to raise capital ratios reduced the ratings of sponsoring banks. Outstanding CP of bank holding companies (almost all directly issued) decreased from a peak of $52 billion in January 1990 to $24 billion at year-end 1991.

The Maturation of Commercial Paper: 1992 to Fall 2007

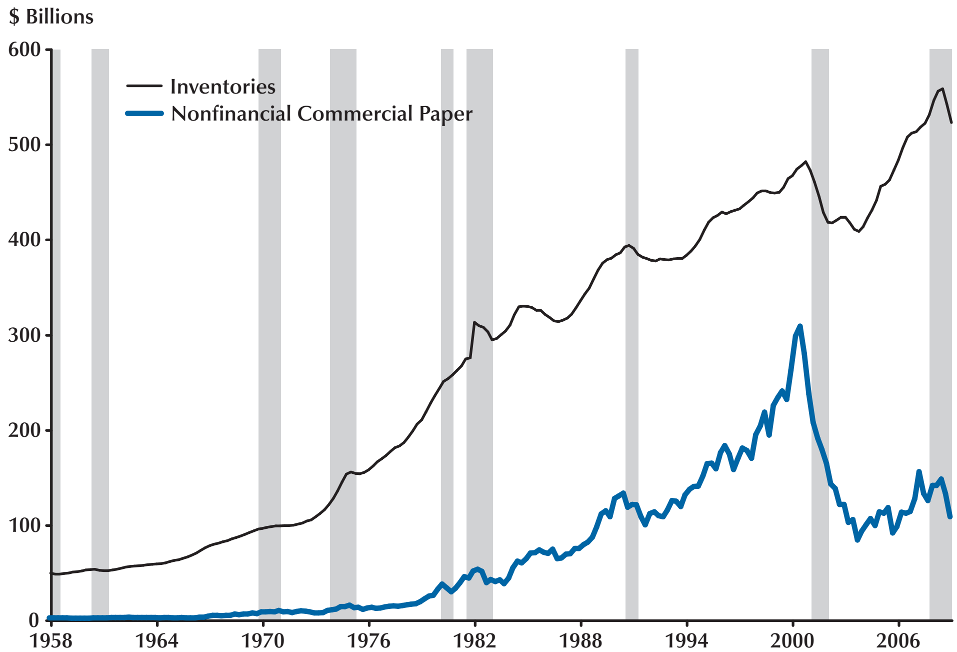
This era of the CP market is characterized by the steady decline in the prominence of nonfinancial CP and the continued rise in ABCP. According to Moody's (2009), ABCP entered the mainstream of money market instruments during the mid-1990s as more institutional investors began to significantly increase their holdings. The ABCP market enlarged in the late 1990s when the commercial bank advisers to ABCP conduits discovered arbitrage opportunities in the securitization of asset-backed securities, residential mortgage-backed securities, and collateralized debt obligations. In general, the arbitrage opportunities arose because the longer-term securities purchased by ABCP conduits carried yields in excess of the London Interbank offering rate (LIBOR), while the conduits could issue short-term (1- to 4-day) ABCP at rates no higher than LIBOR (Standard & Poor's, 2008). Because the rate differential largely reflects the unhedged term premium and the uncovered rollover funding risk, success of the arbitrage depends on the premium not moving sharply.

The Decline in Nonfinancial Commercial Paper. Interaction between the CP market and other types of finance, driven by changes in respective yields, is illustrated by the decrease in nonfinancial CP outstanding during the 2000 recession. Beginning in 2000, total nonfinancial CP outstanding dropped by almost 50 percent in just over 2 years (Figure 6).

Shen (2003) concludes that "aggressive inventory reduction and the widespread practice of replacing [CP] with longer term corporate bonds have reduced the demand for credit in the [CP] market." Because nominal rates were relatively low following the 2000 recession, businesses elected to reduce uncertainty about future borrowing costs by reducing holdings of CP and issuing bonds at low interest rates. Subsequent data have supported Shen's view. The share of nonfinancial businesses borrowing through the CP market declined from 5.4 percent between 1995 and 2000 to 2.3 percent between 2001 and 2008. At the same time, the share of nonfinancial borrow-

Figure 6

Nonfinancial Commercial Paper Outstanding and Manufacturers' Total Inventories



SOURCE: Federal Reserve Board, Census Bureau.

ing through corporate bond issuance increased from 45.9 percent (1995-2000) to 54.4 percent (2001-08).[15]

Innovations in Asset-Backed Commercial Paper. In 2003, additional innovation changed the internal dynamics of ABCP conduits. Previously, typical conduit programs required three players in supporting roles: the adviser, the liquidity enhancer, and the credit enhancer. At times, one commercial or investment bank would play all three roles but, to avoid self-dealing and conflict of interest, the roles typically were played by two or three separate banks. Seeking to increase profits, some bank advisers brought to market ABCP conduits without liquidity and credit enhancers; instead, the advisers asserted that the conduit would rely on its own "internal liquidity" to satisfy all obligations (i.e., securities arbitrage ABCP conduits). Maturing CP that is not rolled over, for example, would be paid off with cash flows generated either from the yields on the assets themselves or by selling the assets (Standard & Poor's, 2008).[16] Nationally recognized statistical rating agencies generally accepted the advisers' assertions but required that such conduits maintain a "cushion" between their ABCP outstanding and the market value of the securities they hold. According to Standard & Poor's (2008)

[15] Shares are calculated using data from the Federal Reserve Flow of Funds table L.2; for example, the share of nonfinancial CP borrowing is nonfinancial CP outstanding divided by nonfederal loans outstanding (net municipal loans, mortgages, and consumer credit loans).

[16] These conduits are commonly referred to as *securities arbitrage* ABCP conduits. Generally speaking, the structures of nonbank investment vehicles, such as hedge funds, closely resemble each other although the terminology differs. For hedge funds, the investment manager handles the portfolio choices, and usually is paid based on performance; the administrator handles back-office tasks including issuing and redeeming shares, calculating net asset value, and measuring fund performance; the prime broker or custodian handles clearing and settlement, money lending, and similar investment banking tasks. In some cases, the roles of administrator and investment adviser are performed by the same firm.

over half of the conduits established between 2004 and 2007 relied, at least in part, on internal liquidity. Acharya, Gale, and Yorulmazer (2009) present a theoretical model of market freezes in which ABCP conduits take a structure very similar to securities arbitrage conduits. The model explains why markets, such as ABCP, that roll over debt can experience sudden freezes.

THE RECENT FINANCIAL CRISIS AND THE COMMERCIAL PAPER MARKET

Financial crises often are defined by sharp increases in the price of risk—that is, the premium that investors require to purchase investments that they previously bought at a much lower yield. CP is a financial instrument particularly susceptible to such an increase in risk premiums. What is not clear, however, is the relative importance of investors' willingness to (i) bear risk and (ii) endure a potential decrease in their liquidity. Clearly, both affect CP market difficulties to some degree. Holders of unsecured traditional CP may suffer significant losses if the issuer fails; in recession, the profit outlooks for most firms dim. In addition, most issuers repay maturing paper by rolling it over; if paper cannot be rolled over and if the issuers' banks do not extend credit to pay the holders, repayment to the holders may be delayed for a considerable period. Layered on top was an increased fear that financial assets, except for U.S. Treasuries, could not be resold to other investors.

The CP market achieved national prominence in the fall of 2008. Heightened financial market uncertainty followed the failure of Lehman Brothers on September 15. Investors and lenders, uncertain of both the creditworthiness of counterparties and their own ability to borrow in the future (if necessary), shortened commitments and shifted away from CP-based products toward default risk-free assets, including MMMFs invested solely in U.S. Treasuries. Borrowers argued that a near closure of the market would sharply worsen the recession. Suddenly, the term "shadow banking system" came into common usage.[17]

Market Events: 2007 and 2008

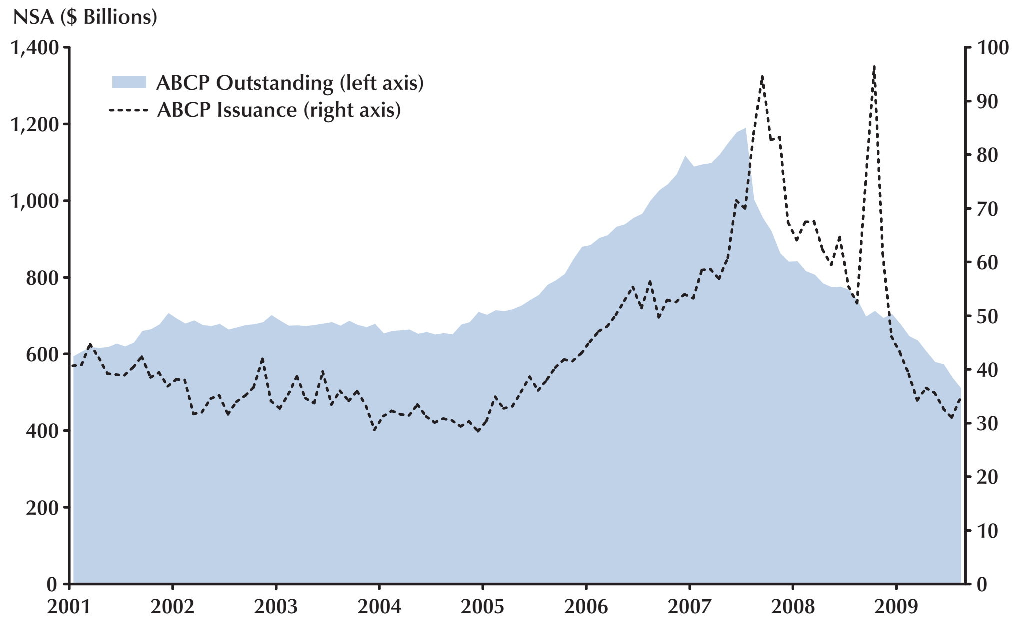
Difficulties in the CP market were apparent during the fall of 2007. Issuance of ABCP, heavily used by mortgage originators to bridge the financing gap between origination and securitization, began to plummet. Mortgage lenders had backed their paper with pools of home loans awaiting securitization. Write-downs on mortgage-related assets caused investors in ABCP to become wary of the underlying assets. A small portion of ABCP issuers (roughly 10 percent) exercised the option allowing them to extend the maturity of their borrowings, thereby cramming longer maturities down to investors expecting repayment (Sahn-Bubna, 2007). In addition, as the market value of residential mortgage-backed securities fell, ABCP conduits relying on internal liquidity began to fail "cushion tests." In some cases, conduits were forced to sell securities—but into a fearful secondary market with few buyers. Between August 6 and 14, 2007, four conduits (representing 1.2 percent of the ABCP market) failed their cushion tests and liquidated their portfolios (Standard & Poor's, 2008). Between August 2007 and July 2008, 27 ABCP conduits with business plans that relied, at least in part, on internal liquidity exited the market (Moody's, 2009).

Figure 7 illustrates the boom and bust in the ABCP market since 2001. Until 2005, the total amount of CP outstanding was relatively stable. Between early 2005 and the summer of 2007, the amount outstanding doubled, reaching a peak of $1.2 trillion in July 2007. As the ABCP market collapsed, some conduits were unable to roll over their paper, resulting in defaults (Keogh, 2007). Investors became increasingly worried that banks, which provided liquidity facilities to the conduits, would be unable to support them (Mollenkamp, 2007). Covitz, Liang, and Suarez (2009) explain how ABCP programs experienced a series of "runs" between August and December 2007. Many runs were directly linked to the credit and liquidity exposures of individual programs. However, the authors provide evidence that the ABCP market was subject to a panic reminiscent of the banking

[17] The term "shadow banking system" refers to those non-bank institutions, such as ABCP conduits, that provide funds to businesses.

Figure 7

Average Monthly ABCP Issuance and Outstanding



NOTE: *Issuance is only AA rated.

SOURCE: Federal Reserve Board, Volume Statistics for Commercial Paper Issuance.

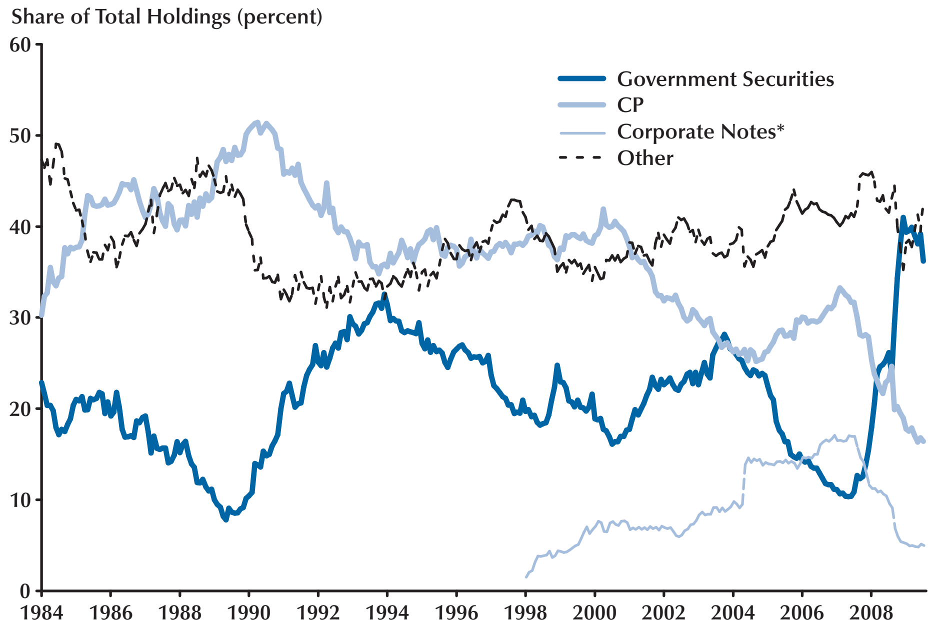
panics during the Great Depression, in which runs on some programs were not even related to program fundamentals. Broad-based investor concerns that sponsoring banks would be unable to meet their commitments if numerous programs required support at the same time caused extensive withdrawals. Startled investors began to shift their holdings from MMMFs invested in ABCP toward MMMFs invested solely in Treasuries (Figure 8).

Difficulties increased during 2008.[18] CP outstanding in December 2008 was $125 billion lower than it was a year earlier, with ABCP paper accounting for half that decrease ($64 billion). At year-end 2008, the amount of outstanding ABCP paper was approximately the same as at year-end 2005. Because much of the intervening increase in ABCP paper was mortgage related, the decrease was not unexpected as the housing market cooled. Moody's (2009) reports that the number of ABCP programs declined to 244 from 265, writing that the ABCP market "is returning to one of primarily bank-sponsored multi-seller programs, much as it was a decade ago." Moody's (2009) reports taking rating actions (that is, reducing or reconsidering ratings) on seven ABCP programs during 2008; in all but one case, the action reflected a weakening of an underlying support party (that is, the liquidity or credit enhancer, usually a bank). Advisers to ABCP conduits struggled to sustain their outstanding issues; one ABCP program defaulted because of decreases in the prices of its assets. On occasion, advisers to ABCP conduits shouldered the responsibility for offsetting asset losses: Nine program advisers declared their intent to financially support their affiliated ABCP programs. Many other advisers provided

[18] This section draws heavily on Moody's Investors Service report, "ABCP 2008 Year in Review and 2009 Outlook," February 10, 2009. For additional details see Fitch Ratings (2008a,b, 2009).

Figure 8

Asset Holdings of Taxable Money Market Mutual Funds



NOTE: *Prior to 1998 Corporate Notes are included in the Other category.

SOURCE: Investment Company Institute.

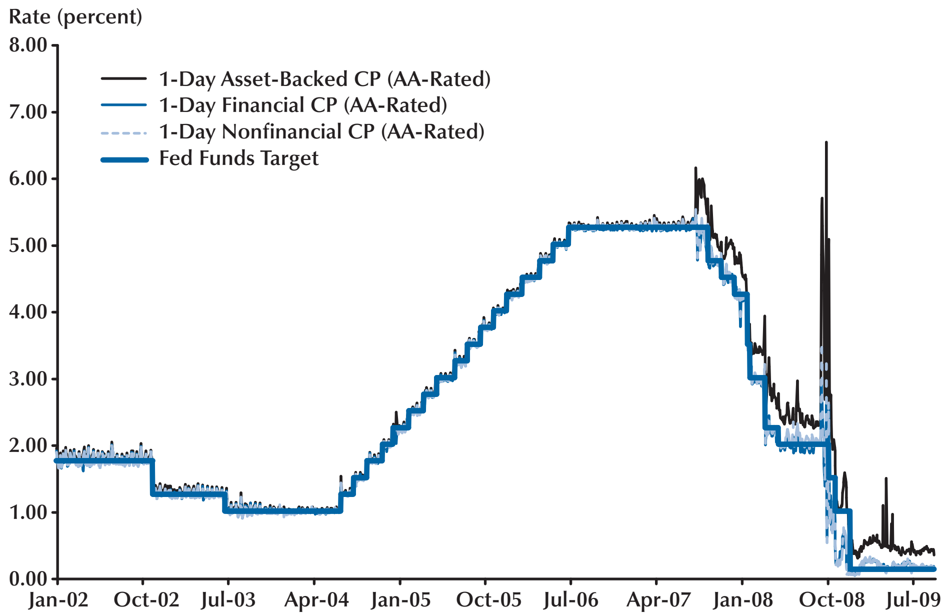
support for individual assets or purchased assets from the conduit to maintain the conduit's credit quality.

The Lehman Brothers bankruptcy on September 15, 2008, was a major disruption to the CP market. During the months prior to bankruptcy, investors had faced a difficult choice: Sell Lehman's paper at a loss on the rumor of failure, or wait and pray for the rescue of Lehman. Lehman's failure brought immediate stress on the CP market. The following day, at 11 a.m., the $62 billion Reserve Primary Fund "broke the buck" (that is, its net asset value fell below $0.995 per nominal share) by writing its Lehman investments (with face value of $785 million) down to 80 cents per share; at 4 p.m., when it wrote the investments down to zero, the fund's net asset value per share reportedly fell to 97 cents and the fund restricted redemptions (Henriques, 2008). Prior to the Reserve Fund actions, it had been 14 years since investors in MMMFs had experienced a loss; in that case, investors were paid 96 cents per share at liquidation.

Investors in institution-type MMMFs, including corporate cash managers, often use the funds in a manner similar to bank deposits and withdraw the funds on short notice. Losses, of course, are undesired—but a suspension of redemptions is intolerable. At that point, the issue became a crisis of liquidity. Reacting to redemption restrictions, investors shifted more than $400 billion from "prime" money funds (invested in CP and other instruments) to money funds invested in Treasuries. Shares in prime funds dropped from $1.3 trillion on September 9 to $864 billion on October 7, while government-only institution-type funds increased by more than $350 billion (Moody's, 2009). The portfolio reallocation in MMMFs is portrayed in Figure 8. In July 2007, approximately 11 percent of MMMFs portfolios were composed of government securities (i.e., Treasury bills); by January 2009, an average of

Figure 9

Selected Overnight Interest Rates



SOURCE: Federal Reserve Board, Table H.15.

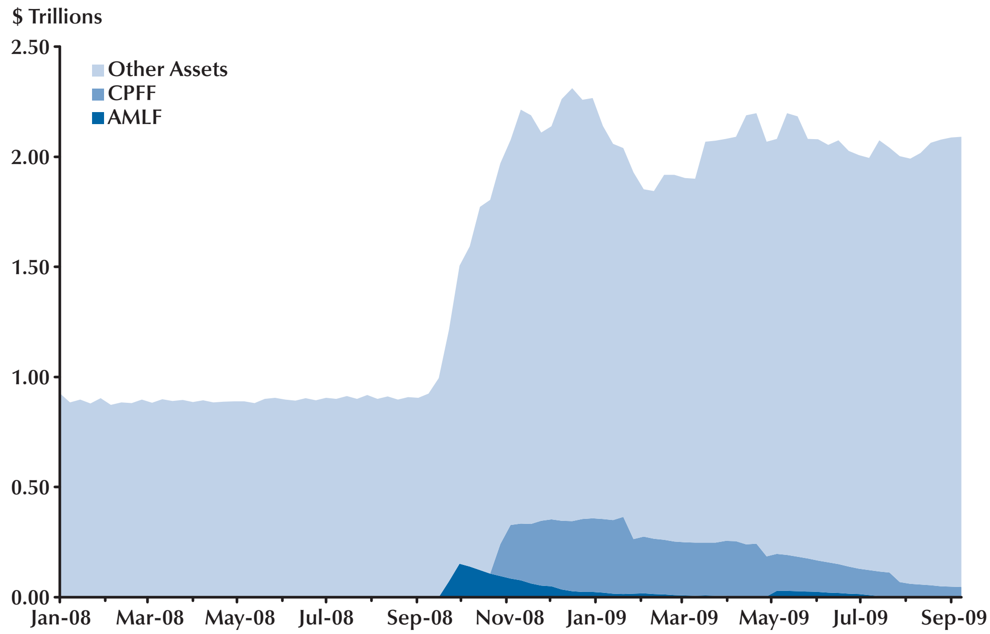
40 percent of a fund's portfolio was made up of government securities. At the same time, the share of CP fell from roughly 32 percent to under 20 percent. Retail-type money funds, held primarily by households but also by smaller businesses, were little affected. Quickly, however, MMMFs that were invested in mortgage-related assets came under pressure. Assistance came from banks and fund managers—published reports said more than $10 billion was pledged. Nevertheless, the demand for CP fell. Issuance dropped, and brokers and dealers were forced to retain elevated inventories; at the end of 2008's third quarter, dealers held $154 billion of ABCP for sale, 78 percent more than a year earlier.[19] It became difficult to place ABCP for terms longer than overnight; overnight issuance increased from approximately 60 percent of the total to as much as 90 percent. Federal Reserve Board data show, however, that total issuance decreased little—the effect of heightened uncertainty was reflected in investors' unwillingness to commit liquid funds for more than one day at a time. Hence, the crisis was primarily one of liquidity—"If I lend today but need to borrow tomorrow, will anyone then lend to me?"—rather than of heightened default risk.

The degree of stress in the ABCP market is reflected in short-term funding rates, all of which increased sharply mid-September, both in absolute level and relative to overnight federal funds (Figure 9). In normal times, CP yields (especially on ABCP) only slightly exceed those on comparable Treasuries. Yet, two prominent spikes are evident in the ABCP rate: late-August to early-September 2007, when mortgage-related write-downs began, and the most prominent in September 2008. A large factor in the September 2008 spike was the scramble by CP issuers for funds when MMMF demand for CP collapsed (Moody's, 2009).

[19] These increases were short-lived. By the fourth quarter of 2008, brokers in dealers were able to shed roughly 60 percent of their holdings.

Figure 10

Federal Reserve Assets (EOP, Wednesday)



SOURCE: Federal Reserve Board, H.4.1 Table 9.

Recent Treasury and Federal Reserve Programs

Following mid-September 2008 market disruptions, the Treasury and Federal Reserve introduced programs to enhance liquidity in two ways: (i) by reducing extension risk, that is, the risk that an investor will not repay maturing CP in a timely fashion, either by rolling the paper or bank borrowing; and (ii) by reducing the risk of suspension of redemptions at MMMFs that hold CP. The Treasury, in an effort to assure investors that future suspension of redemptions would not occur, offered insurance for the value of MMMF shares held as of September 18 at funds choosing to participate in its program.[20] The Federal Reserve introduced three programs with varied objectives, including assuring money fund managers that their CP could be sold quickly if necessary and providing a funding facility that issuers of highly rated paper could use as a backstop if rollover failed.

The assets on the Fed's balance sheet are shown in Figure 10. In response to the financial crisis the Fed created numerous lending programs, causing its total assets to increase from under $1 trillion to over $2 trillion.[21] The two dark-blue areas represent the assets held by two programs focused on the CP market: the Commercial Paper Funding Facility (CPFF) and the Asset-Backed Commercial Paper Money Market Fund Liquidity Facility (AMLF). Roughly 15 percent of the Fed's assets were acquired through these programs. By comparison, less than 1 percent of the Fed's assets were acquired from Bear Stearns or loans to American International Group (AIG). We review each of these programs below.

Money Market Investor Funding Facility. This program was authorized by the Federal

[20] See U.S. Department of the Treasury (2008) for additional details.

[21] See Gavin (2009) and Gascon (2009) for discussion of other Fed lending programs and their impact on the Fed's balance sheet.

Reserve Board on October 21, 2008, under the provisions of Section 13(3) of the Federal Reserve Act.[22] The Board describes the program as allowing the Federal Reserve Bank of New York to provide credit to "a series of special purpose vehicles" established "by the private sector" to purchase from eligible investors "certain highly rated, short-term instruments, including certificates of deposit, bank notes, and CP." Essentially, the SPVs would be authorized to purchase bank debt or CP forcibly sold into the market as a result of a run on a bank or money fund. This program had no activity and expired on October 30, 2009.

Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility. Created on September 19, 2008, the AMLF essentially allows money market funds indirect access to the Federal Reserve discount window via a depository financial institution. The Federal Reserve Board describes this facility as follows:

> [AMLF] is a lending facility that provides funding to U.S. depository institutions and bank holding companies to finance their purchases of high-quality...ABCP from money market mutual funds under certain conditions. The program is intended to assist money funds that hold such paper in meeting demands for redemptions by investors and to foster liquidity in the ABCP market and money markets more generally.

Because money funds themselves are not eligible to borrow at the discount window, to borrow under AMLF they first must sell ABCP to an eligible depository institution. Similar to other programs that seek to assure investors that a suspension of redemptions will not occur in the future, only ABCP owned prior to the AMLF's inception is eligible. Assets of the AMLF initially expanded rapidly, reaching a maximum of $152 billion in its second week (the AMLF has the option to resell the paper or hold it until maturity). Since that time, the program's assets have decreased at an average weekly rate of 11 percent. In the week ending April 1, 2009, the AMLF held just over $6 billion in assets.

On June 25, 2009, the authorization of the AMLF was extended through February 1, 2010, although with additional administrative criteria to ensure the program is used for its intended purpose of a temporary liquidity backstop.

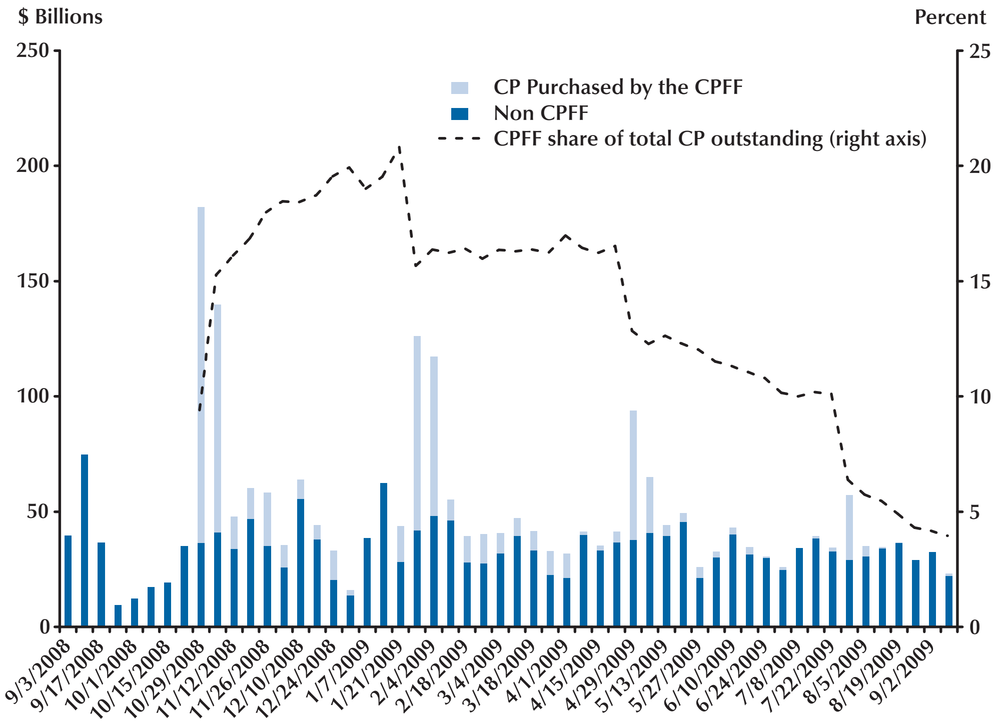
Commercial Paper Funding Facility. The events of mid-September 2008 made money market investors (who prize liquidity) hesitant to purchase assets with maturities longer than a single day. In normal times, approximately 5 to 10 percent of daily CP issuance is 91-day maturity, and represents 20 to 25 percent of all outstanding paper. In mid-September, 91-day issuance fell to near zero. On Friday, September 12, for example, 60 percent of issuance was 1- to 4-day maturity; by Wednesday, September 17, 87 percent was 1- to 4-day maturity. On October 7, 2008, the Federal Reserve announced the creation of the CPFF to support longer-maturity paper. The CPFF's structure is similar to the Money Market Investor Funding Facility: An SPV purchases 3-month corporate unsecured and asset-backed A1/P1–rated CP using funds provided by the Federal Reserve Bank of New York.[23] The paper is held to maturity. Similar to other CP market support programs, the program is linked to the events of mid-September 2008: The maximum amount an issuer can sell to the CPFF is the maximum amount the issuer had outstanding between January 1 and August 31, 2008, and the CPFF will not purchase from issuers who were inactive prior to its inception. The first purchases by the CPFF occurred on October 27, 2008. It was originally scheduled to purchase paper through on October 30, 2009, but was extended through February 1, 2010, in order to ensure the access of U.S. businesses to short-term funding. However, the interest rates of the CPFF have become increasingly unattractive to many borrowers.

[22] Section 13(3) allows the Federal Reserve Banks, under certain conditions and with specified approval of the Board of Governors, to lend to almost any borrower via the discounting of assets. Section 13(3) does not permit direct lending; rather, the funding is supplied via the borrower discounting assets to the Federal Reserve. Hence, using a SPV as the borrower is convenient.

[23] The details are more complex than summarized here. Technically, the facility purchases newly issued paper with maturity of 81 days or more. Pricing also is complex, with surcharges of 100 to 300 basis points. Each participating company must also pay a registration fee to use the CPFF. For details, see Federal Reserve Bank of New York (2009b).

Figure 11

CPFF Share of Purchases and Outstanding



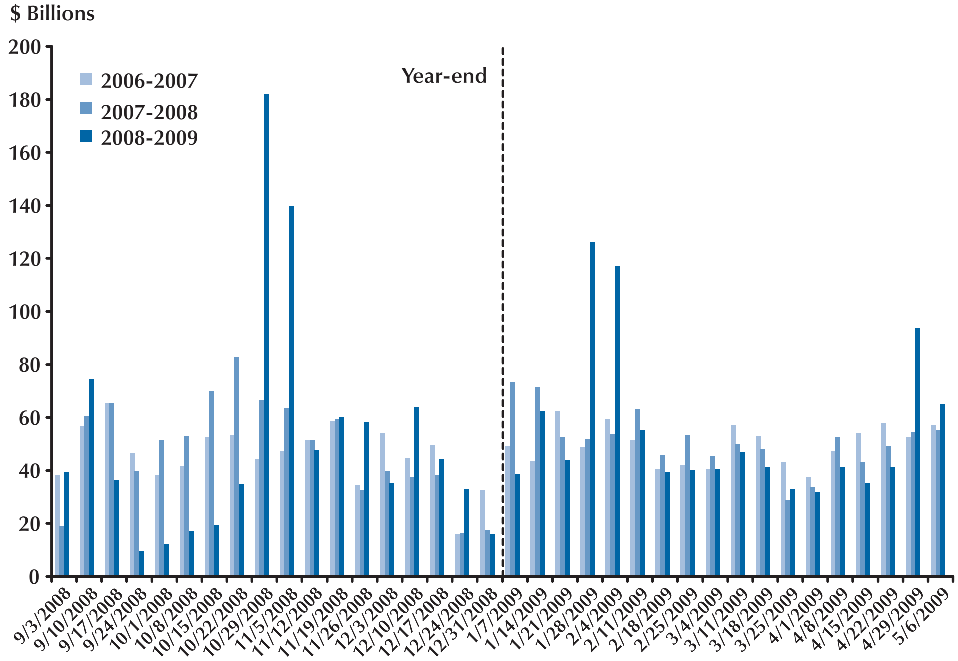
SOURCE: FRB, H.4.1 Table 1; Federal Reserve Bank of New York.

The CPFF has been the most active of the Federal Reserve's three support programs for the CP market and has been cited by Chairman Bernanke (2009) and others as a highly successful market support activity. Hence, its history is worthy of closer examination.

Figure 11 shows the weekly issuance of 3-month CP (rated A1/P1 and A2/P2), between September 2008 and February 2009. Issuance decreased sharply during September 2008, but increased steadily during October. During its first two weeks, the CPFF purchased the overwhelming majority of all newly issued eligible 3-month CP. One likely reason for such large volume was the wish by corporations to lock in year-end financing; daily data show that issuance jumped on October 27-29, the first days of purchase by the CPFF. For all weeks thereafter—until the week of January 28, 2009, when the initially purchased 91-day paper matured—relatively little CP was purchased by the CPFF. The second burst of CPFF activity occurred the weeks of January 28 and February 4, when paper purchased by the CPFF in October rolled over. Later weeks show light activity. When the CPFF was in full swing, it held over 20 percent of all CP outstanding, but fewer and fewer investors continued to roll over their paper with the CPFF; at the time of this writing, the CPFF currently holds less than 5 percent of all CP outstanding. The reason likely reflects a pricing policy designed to urge private sector, not CPFF, funding. Funding via the CPFF is not inexpensive, with pricing set to yield 100 to 300 basis points above the overnight index swap rate. Further, some former CPFF borrowers have turned to the Temporary Liquidity Guarantee Program of the Federal Deposit Insurance Corporation, which guarantees bank debt at far longer maturities.

Figure 12

Commercial Paper Issuance



SOURCE: Federal Reserve Board, Volume Statistics for Commercial Paper Issuance.

Figure 12 shows some historical perspective, which compares, side by side, weekly issuance in 2006-07, 2007-08, and 2008-09 of CP with maturity greater than 80 days. The reduction in CP issuance near year-end is quite pronounced.[24] Although the period mid-September to mid-October 2008 clearly is unusual, issuance appears largely to have recovered by the time of the CPFF's first purchases. Was the CPFF necessary? Would the market have recovered in the absence of the CPFF? Or was the CPFF's presence essential to assure investors that a "purchaser of last resort," similar to the Federal Reserve's discount window, was available to mitigate rollover risk? It is too early to say, as of this writing.

Finally, we note that the CPFF program has been profitable for the Federal Reserve. According to recently issued financial statements (Federal Reserve Bank of New York, 2009a), between October 14 and December 31, 2008, the program had a net income of $1.08 billion and, as of December 31, 2008, the program had experienced no defaults.

CONCLUSION: THE FUTURE OF COMMERCIAL PAPER

The CP market and MMMFs have matured together, each complementing the other, and today are the liquid core of the U.S. shadow banking system. Money funds intermediate CP into liquid shares that have many of the characteristics of bank deposits; that is, the money funds provide investors—large or small, retail or institutional—a liquid, high-quality, low-risk investment alternative. Simultaneously, money funds purchase CP.

The CP market was originated by firms seeking short-term funds at interest rates and terms

[24] Musto (1997) and Downing and Oliner (2007) provide a discussion about the year-end effects in the CP market.

more favorable than bank loans. The rise of bank-advised, multiseller ABCP conduits during the 1990s extended the market's purpose such that it became focused on asset securitization and risk diffusion, typically with significant off-balance-sheet support from the nation's largest commercial banks. Today, the economic role of conduits is similar to the role played by banks. Assets purchased by conduits provide funds to businesses small and large, while conduits' bank advisers seek to monitor the management and performance of those assets' issuers. Simultaneously, the CP issued by conduits provides to investors a liquid, low-risk asset. All of this occurs without the cost and fuss of a banking charter, capital adequacy requirements, or federal deposit insurance.

The rapid growth and large size of the CP market sensitizes it to adverse events, including the bankruptcy of the Penn Central Railroad in 1970 and, more recently, the bankruptcy of Lehman Brothers. The sensitivity is twofold: Issuers of secured paper find it increasingly difficult to roll over their paper, even at shorter maturity and higher cost, and MMMFs and ABCP programs may experience runs. The causes and consequences of these two sensitivities require further research.[25] Policymakers will also find it necessary to address if and how the regulations will need to be implemented in what is now understood to be a systematically important sector of the U.S. and global economy.[26] On the other hand, in a low-interest-rate environment, businesses may prefer to secure long-term financing and shift away from CP. To the extent that financial markets currently expect low interest rates to prevail for an extended period, volume in the CP market may be attenuated for some time.

[25] We refer readers to Acharya, Gale, and Yorulmazer (2009) for starting points for the discussion on rollover risk, and Covitz, Liang, and Suarez (2009) on runs of ABCP programs.

[26] Acharya, Gale, and Yorulmazer (2009) suggest improving the liquidation value of assets and higher capital requirements as possible solutions. Gatev and Strahan's (2006) results suggest that commercial banks should naturally be well positioned to act as liquidity providers during crises as a "flight to quality" will boost bank reserves, allowing them to meet the demands of their ABCP programs. More obvious remedies would be the permanent establishment of a CPFF-type backstop.

REFERENCES

Acharya, Viral V.; Gale, Douglas M. and Yorulmazer, Tanju. "Rollover Risk and Market Freezes." Federal Reserve Bank of New York Working paper, February 2009.

Anderson, Richard G. "Bankers' Acceptances: Yesterday's Instrument to Restart Today's Credit Market?" Federal Reserve Bank of St. Louis *Economic Synopses*, 2009a, No. 5, January 9; http://research.stlouisfed.org/publications/es/09/ES0905.pdf.

Anderson, Richard G. "Bankers' Acceptances and Unconventional Monetary Policy: FAQs." Federal Reserve Bank of St. Louis *Economic Synopses*, 2009b, No. 14, March 18, 2009; http://research.stlouisfed.org/publications/es/09/ES0914.pdf.

Board of Governors of the Federal Reserve System. *Banking and Monetary Statistics 1941-1970*. Washington, DC: Board of Governors, 1976; http://fraser.stlouisfed.org/publications/bms2/.

Bernanke, Ben S. "The Crisis and the Policy Response." Stamp Lecture, London School of Economics, January 13, 2009; www.federalreserve.gov/newsevents/speech/bernanke20090113a.htm.

Calomiris, Charles W. "Is the Discount Window Necessary? A Penn Central Perspective." Federal Reserve Bank of St. Louis *Review*, May/June 1994, *76*(3), pp. 31-55.

Cantor, Richard and Rodriques, Anthony P. "Nonbank Lenders and Credit Slowdown," in *Studies on Causes and Consequences of the 1989-92 Credit Slowdown*. New York: Federal Reserve Bank of New York, February 1994.

Covitz, Daniel; Liang, Nellie and Suarez, Gustavo. "The Evolution of a Financial Crisis: Panic in the Asset Backed Commercial Paper Market." Federal Reserve Board Finance and Economics Discussion Series: 2009-36, August 18, 2009; www.federalreserve.gov/pubs/feds/2009/200936/index.html.

Downing, Chris and Oliner, Stephen. "The Term Structure of Commercial Paper Rates." *Journal of Financial Economics*, January 2007, *83*(1), pp. 59-86.

Federal Reserve Bank of New York. "Commercial Paper Funding Facility LLC: Financial Statements for the Period October 14, 2008 to December 31, 2008, and Independent Auditors Report." April 2, 2009a; www.newyorkfed.org/aboutthefed/annual/annual08/CPFFfinstmt2009.pdf.

Federal Reserve Bank of New York. "Commercial Paper Funding Facility: Program Terms and Conditions" (effective June 25, 2009b); www.newyorkfed.org/markets/cpff_terms.html.

Fitch Ratings. "Global Asset-Backed Commercial Paper: 2007 Review and 2008 Outlook." *Structured Finance: Asset-Backed Special Report*, January 24, 2008.

Fitch Ratings. "The ABCP Paper Trail." May 2008.

Fitch Ratings. "2009 Outlook for Global ABCP." *Structured Finance: Asset-Backed Special Report*, February 12, 2009.

Gascon, Charles S. "Federal Reserve Assets: Understanding the Pieces of the Pie." Federal Reserve Bank of St. Louis *Economic Synopses*, 2009, No. 13; March 10, 2009; http://research.stlouisfed.org/publications/es/09/ES0913.pdf.

Gatev, Evan and Strahan, Philip E. "Banks' advantage in hedging liquidity risk: theory and evidence from the commercial paper market." *Journal of Finance*, April 2006, *62*(2), 867-91.

Gavin, William T. "More Money: Understanding Recent Changes in the Monetary Base." Federal Reserve Bank of St. Louis *Review*, March/April 2009, *91*(2), pp. 49-59; http://research.stlouisfed.org/publications/review/09/03/Gavin.pdf.

Hahn, Thomas K. "Commercial Paper" in Timothy Q. Cook and Robert K. Laroche, eds., *Instruments of the Money Market*. Federal Reserve Bank of Richmond Special Report, 1998, pp. 105-27; www.richmondfed.org/publications/research/special_reports/instruments_of_the_money_market/pdf/chapter_09.pdf.

Henriques, Diana B. "Buck Broken, But Timing May Affect Redemptions." *New York Times*, November 26, 2008. www.nytimes.com/2008/11/27/business/27fund.html?_r=1&scp=1&sq=Henriques%20and%20November%2027,%202008%20and%20Buck%20broken&st=cse.

Kavanagh, Barbara; Boemio, Thomas R. and Edwards, Gerald A. "Asset-Backed Commercial Paper Programs." *Federal Reserve Bulletin*, February 1992, *78*(2), pp. 107-16.

Keogh, Bryan. "Asset-Backed Commercial Paper Drops Most in 2 Months," Bloomberg, November 8, 2007.

Mollenkamp, Carrick. "Shake-Up At Citigroup—Credit Crunch: Commercial-Paper Market Hit by New Investor Anxiety" *Wall Street Journal*, November 5, 2007.

Moody's Investor Service. "The Fundamentals of Asset-Backed Commercial Paper" *Structured Finance: Special Report*, Moody's Investors Service, February 3, 2003.

Moody's Investor Service, "ABCP 2008 Year In Review and 2009 Outlook" *Structured Finance: Special Report*, Moody's Investors Service, February 10, 2009.

Musto, David K., "Portfolio Disclosures and Year-End Price Shifts," *Journal of Finance*, September 1997, *52*(4), pp. 1563-88.

Post, Mitchell A. "The Evolution of the U.S. Commercial Paper Marker Since 1980." *Federal Reserve Bulletin*, December 1992, *78*(12), pp. 880-91.

Sahn-Bubna, Aparajita. "Commercial Paper Shows Some Stress—Bad Mortgages Weigh On Low-Profile Corner of Short-Term Market." *Wall Street Journal*, August 8, 2007.

Shen, Pu. "Why Has the Nonfinancial Commercial Paper Market Shrunk Recently?" Federal Reserve Bank of Kansas City, *Economic Review*, First Quarter 2003.

Standard and Poor's. "The U.S. Asset-Backed Commercial Paper Market May be Down, But It's Not Out." Standard and Poor's *Ratings Direct*, July 7, 2008.

Stigum, Marcia and Crezcenzi, Anthony. *Stigum's Money Market*. Fourth edition. New York: McGraw-Hill, 2007.

United States Department of the Treasury. "Frequently Asked Questions About Treasury's Temporary Guarantee Program for Money Market Funds." Press Release: HP-1163, September 29, 2008.

Local Price Variation and Labor Supply Behavior

Dan A. Black, Natalia A. Kolesnikova, and Lowell J. Taylor

In standard economic theory, labor supply decisions depend on the complete set of prices: wages and the prices of relevant consumption goods. Nonetheless, most theoretical and empirical work in labor supply studies ignore prices other than wages. We address the question of whether the common practice of ignoring local price variation in labor supply studies is as innocuous as generally assumed. We describe a simple model to demonstrate that the effects of wage and nonlabor income on labor supply typically differ by location. In particular, we show that the derivative of the labor supply with respect to nonlabor income is independent of price only when the labor supply takes a form based on an implausible separability condition. Empirical evidence demonstrates that the effect of price on labor supply is not a simple "up-or-down shift" that would be required to meet the separability condition in our key proposition. (JEL J01, J21, R23)

Federal Reserve Bank of St. Louis *Review*, 2009, *91*(6), pp. 613-25.

In standard economic theory, labor supply decisions depend on the complete set of prices: the wages and the prices of relevant consumption goods. Nonetheless, as Abbott and Ashenfelter (1976) noted some 30 years ago, economists generally have found it a useful abstraction, in both theoretical and empirical work, to ignore prices other than wages in labor supply studies. For example, none of the empirical results on labor supply discussed in the prominent reviews of Pencavel (1986), Killingsworth and Heckman (1986), or Blundell and MaCurdy (1999) are derived by procedures that account for variation in any price other than wages.[1]

However, most empirical work on labor does use national datasets of individuals who live in different locations and therefore face different prices for locally priced goods. These price differences can be quite large, especially for housing. For example, according to 1990 Census data, the median housing price in New York is more than three times that of the median housing price in Cleveland.[2] The question addressed in this paper is whether the common practice of ignoring local price variation in labor supply studies is as innocuous as has generally been assumed.

To examine the issue, we first present a simple theoretical model: an economy in which people live in different locations with differing levels of a production or consumption amenity. Following

[1] Abbott and Ashenfelter's (1976) evaluation of labor supply in the United States for the 1929-67 period exploits time-series changes in relative prices but does not evaluate possible impacts of cross-sectional variation (which, as they state, is "expected to be small"). Some work conducts sensitivity analysis using Bureau of Labor Statistics information on the cost of living to "adjust" wages. See, for instance, DaVanzo, DeTray, and Greenberg (1973) and Masters and Garfinkel (1977).

[2] Gabriel and Rosenthal (2004) and Chen and Rosenthal (2008) show that massive housing price differences pertain across cities even after careful adjustment for quality.

Dan A. Black is a professor in the Harris School, University of Chicago, and a senior fellow at the National Opinion Research Center; Natalia A. Kolesnikova is an economist at the Federal Reserve Bank of St. Louis; and Lowell J. Taylor is a professor of economics at the Heinz College, Carnegie Mellon University. An earlier version of this article was published in the Federal Reserve Bank of St. Louis *Regional Economic Development*, 2008, *4*(1), pp. 2-14.

logic familiar in urban economics, (e.g., Roback, 1982), equilibrium prices will differ across locations. We demonstrate that labor supply behavior also can vary across locations.

Next, we demonstrate that, when prices vary across locations, local variation in prices can be safely ignored only when preferences take a very specific and peculiar form. We also show that the responsiveness of labor supply to *wage* changes will be the same across locations only if the responsiveness of labor supply to *nonlabor income* changes is the same across locations.

In our third step we evaluate the potential empirical importance of our theoretical observations. We present results obtained by using 1990 Public Use Microdata Samples (PUMS) of the 1990 U.S. Census that examine labor supply in the nation's 50 largest cities. We focus on the labor force participation and hours decisions of white married women aged 30 to 50—a group whose labor decisions are quite responsive to changes in wages and nonlabor income.

In general, we analyze the basic "building block" empirical relationship that would underlie any empirical analysis of labor supply for this group: the relationship between nonlabor income and labor supply. Our innovation is examining this relationship for each of the 50 cities separately and demonstrating the significant systematic variation that exists among them.

We find that the basic correlation—between labor supply and nonlabor income—differs across cities. For example, women who have relatively high nonlabor income (primarily a husband's income) work relatively fewer hours and have lower participation rates. An important observation, from our perspective, is that this anticipated negative relationship is substantially more pronounced in cities with inexpensive housing than in cities with expensive housing.

A MODEL OF LOCAL LABOR MARKETS WITH STONE-GEARY PREFERENCES

We begin our study by presenting a simple model of local price variation along the lines of Roback (1982) and Haurin (1980). Locations differ based on two criteria: (i) A location may be inherently more pleasant (i.e, have a higher level of a "consumption amenity," such as nice weather), or (ii) a location may be associated with inherently higher productivity (e.g., owing to the presence of a natural resource or an agglomeration of economies in production). For simplicity we restrict attention to cases in which people choose to live in one of two cities.

In contrast to the standard urban location models such as those of Roback (1982) or Haurin (1980), which fix labor supply as a constant, we allow labor supply to be a choice variable. Preferences are assumed to be Stone-Geary. This is a particularly transparent form of utility, and as Ashenfelter and Ham (1979) note, it is the simplest functional form of utility used in applied empirical work examining labor supply.[3] We assume, in particular, that individual i has utility u^i as a function of a consumption good x, leisure l (which is scaled so that $0 \le l \le 1$), and an amenity level A^j (that is specific to location j), according to a simple Stone-Geary form as follows:

$$u^i = \theta^{ij} A^j \left(x - c\right)^{\delta} l^{1-\delta}, \tag{1}$$

where c and δ are parameters that are common across individuals and θ^{ij} is a positive idiosyncratic parameter that equals 1 for a typical individual, but allows for the possibility that person i has a particular attraction, or distaste, for location j (as θ^{ij} is greater than, or less than, 1).

A person living in location j maximizes utility subject to a budget constraint, $p_j x = w_j(1 - l) + N$, where p_j is the price for the local consumption good, w_j is the local wage, and N is nonlabor income. Assuming an interior solution pertains, demand for leisure and for the consumption good are, respectively,

$$l\left(w_j, p_j\right) = \frac{\left(1-\delta\right)\left(N + w_j - cp_j\right)}{w_j}, \tag{2}$$

and

[3] See also Blundell and MaCurdy (1999) for a discussion of the Stone-Geary form, as well as other forms used in applied work on labor supply.

$$(3) \qquad x(w_j, p_j) = \frac{\delta(N + w_j - cp_j)}{p_j} + c.$$

Substituting equations (2) and (3) into equation (1) gives indirect utility for person in location

$$(4) \qquad V^{ij} = \frac{\theta^{ij} A^j \delta^\delta (1-\delta)^{1-\delta} (N + w_j - cp_j)}{p_j^\delta w_j^{1-\delta}}.$$

In equilibrium each individual chooses to live in the location that yields the highest level of utility. There are two locations: $j = 1$ or 2. We present two cases: one with differing consumption amenities and one with differing levels of productivity in the locations.

Case 1: Differing Levels of the Consumption Amenity

Suppose there is general agreement that Location 1 is nicer than Location 2, $A^1 > A^2$, and for the moment assume further that there are no idiosyncratic differences in opinion about location, so that $\theta^{ij} = 1$ for all individuals. Because workers are equally productive in the two locations, wages and w_1 and w_2 must be the same, say w.[4] In an equilibrium in which people live in both locations, we must have $V^{i1} = V^{i2}$, so using equation (4), it is clear that p_1 and p_2 must solve

$$(5) \qquad \frac{A^1(N + w - cp_1)}{p_1^\delta w^{1-\delta}} = \frac{A^2(N + w - cp_2)}{p_2^\delta w^{1-\delta}}.$$

Inspection of equation (5) confirms the intuitive result that $p_1 > p_2$: The local consumption good is more expensive in Location 1—the high-amenity city.

This logic continues to hold if we add back the idiosyncratic taste component to utility. If for the marginal individual $\theta^{i1} = \theta^{i2} = 1$, equation (5) still characterizes equilibrium prices. In this instance, however, some individuals will have a strict preference with regard to location. For example, an individual with $\theta^{i1} > \theta^{i2}$ will have a strict preference for Location 1 over Location 2.

We turn next to labor supply. Let h be the fraction of time that a person works, $h = 1 - l$. From equation (2), we have

$$(6) \qquad h(w, p_j) = \frac{\delta w - (1-\delta)(N - cp_j)}{w}.$$

Although wages are the same in both locations, the labor supply differs. In this example, $h(w,p_1) > h(w,p_2)$; individuals supply more labor when they work in the more expensive city.

Suppose instead the focus is on the effect of a wage change in a local labor market (studying people who would not move in response to a small change in the wage)[5]:

$$(7) \qquad \frac{\partial h(w, p_j)}{\partial w} = \frac{(1-\delta)(N - cp_j)}{w^2}.$$

Notice that in this example, the responsiveness of the labor supply to a wage change is greater in the inexpensive city than in the expensive city,

$$\frac{\partial h(w, p_2)}{\partial w} > \frac{\partial h(w, p_1)}{\partial w}.$$

In contrast, if we focus on how a change in nonlabor income affects labor supply,

$$(8) \qquad \frac{\partial h(w, p_j)}{\partial N} = \frac{-(1-\delta)}{w},$$

we find that the relationship is independent of the local price; that is, it can be written as

$$\frac{\partial h(w)}{\partial N}.$$

Case 2: Differing Levels of Productivity

Now suppose that Locations 1 and 2 are viewed as equally pleasant, $A^1 = A^2$, but productivity is higher in Location 1 than in Location 2,

[4] For simplicity, we are implicitly assuming that labor is the only factor of production, so that firms will be indifferent in hiring if the wage is the same in the two cities. This would not be true, for example, if land were a major factor of production and land prices differed in the two cities.

[5] In general, if the wage increases in a labor market, this factor can attract new individuals to that location. Here, we are interested in the effect on the labor supply of individuals who are already in the market, for example, people who have an idiosyncratic taste for that location.

so that $w_1 > w_2$. The equilibrium condition corresponding to equation (5)—that the marginal individual is indifferent between locations (i.e., $V^{i1} = V^{i2}$)—is then

$$(9)\qquad \frac{\left(N+w_1-cp_1\right)}{p_1^{\delta}w_1^{1-\delta}}=\frac{\left(N+w_2-cp_2\right)}{p_2^{\delta}w_2^{1-\delta}}.$$

As for labor supply, in city *j*,

$$(10)\qquad h\left(w_j,p_j\right)=\frac{\delta w_j-\left(1-\delta\right)\left(N-cp_j\right)}{w_j}.$$

In general, labor supply differs in the two locations, but even with $p_1 > p_2$ and $w_1 > w_2$ the location that will have the larger labor supply cannot be predicted. Similarly, in general

$$\frac{\partial h\left(w_1,p_1\right)}{\partial w}\neq\frac{\partial h\left(w_2,p_2\right)}{\partial w},$$

and we cannot determine in which city the labor supply is more responsive to wage changes. On the other hand, in this example the derivative of labor supply with respect to nonlabor income,

$$(11)\qquad \frac{\partial h\left(w_j,p_j\right)}{\partial N}=\frac{-\left(1-\delta\right)}{w_j},$$

turns out to be independent of p_j. Furthermore, the derivative of labor supply with respect to nonlabor income does not depend on the local price, p, but because in equilibrium the high-productivity city has relatively higher wages, we expect to observe that $\delta h/\delta N$ will be smaller (in absolute value) in the expensive city.

Our examples illustrate two important points. First, cross-sectional variation in wages and prices may be associated with variation in labor supply, although that cross-sectional variation is of no value for understanding the behavioral effect of wage changes on labor supply. For instance, in our Case 2, even if in both cities

$$\frac{\partial h\left(w_j,p_j\right)}{\partial w}>0,$$

identical individuals may well supply *less* labor in the high-wage city than in the low-wage city, depending on the local price-wage combination. Second, the responsiveness of labor supply to changes in the wage or nonlabor income typically varies across locations.

WHEN DOES PRICE VARIATION MATTER FOR LOCAL LABOR SUPPLY?

As noted previously, housing prices vary widely across U.S. cities, presumably because of differences in consumption or production amenities across these locations. The examples in the previous section indicate that labor supply varies across locations even in the unusually simple and transparent case of Stone-Geary preferences. We now turn to a more systematic investigation of conditions on preferences under which price and income effects on labor supply do *not* depend on location. As is common in the literature, attention is restricted to the case of quasi-homothetic preferences (of which Stone-Geary is a special case).[6] Given this common simplification, what further restrictions are necessary to allow investigators to ignore variation across locations when examining labor supply?[7]

Under quasi-homothetic preferences, indirect utility takes the form

$$(12)\qquad V\left(p,w,N\right)=\alpha\left(p,w\right)+\left(N+w\right)\beta\left(p,w\right),$$

where, as before, p is the local price, w is the local wage, and N is the nonlabor income. Using Roy's identity we derive the demand for leisure

[6] Quasi-homothetic preferences are useful because they preserve a linear expansion path of homothetic preferences, but they do not require the path to go through the origin. Thus, under quasi-homothetic preferences, income elasticities of demand need not equal 1, as is the case with homothetic preferences.

[7] We could attempt to analyze cases that are even more general, but as we shall see, matters are sufficiently discouraging even for the quasi-homothetic case.

$$
\begin{aligned}
l(p,w,N)-1 &= -\frac{\partial V/\partial w}{\partial V/\partial N} \\
&= -\frac{\alpha_w(p,w)+\beta(p,w)+(N+w)\beta_w(p,w)}{\beta(p,w)} \\
&= -\frac{\alpha_w(p,w)+(N+w)\beta_w(p,w)}{\beta(p,w)}-1, \\
l(p,w,N) &= -\frac{\alpha_w(p,w)+(N+w)\beta_w(p,w)}{\beta(p,w)}.
\end{aligned} \tag{13}
$$

It then follows that hours of labor supply are

$$
\begin{aligned}
h(p,w,N) &= 1-l(p,w,N) \\
&= 1+\frac{\alpha_w(p,w)+(N+w)\beta_w(p,w)}{\beta(p,w)}: \\
&= a(p,w)+(N+w)b(p,w),
\end{aligned} \tag{14}
$$

where $a(p,w)=1+\frac{\alpha_w}{\beta},\ b(p,w)=\frac{\beta_w}{\beta}$.

Consider the effect of the change in nonlabor income on the labor supply,

$$\frac{\partial h}{\partial N}=b(p,w)=\frac{\beta_w(p,w)}{\beta(p,w)}.$$

Obviously, $\delta h/\delta N$ is independent of p (and thus is the same across locations) if and only if $b(p,w) \equiv b(w)$. The following claim provides the condition under which this holds:

Claim $\frac{\beta_w(p,w)}{\beta(p,w)}=b(w)\Leftrightarrow \beta(p,w)=\beta_1(p)\beta_2(w)$.

Proof. The proof of sufficiency is trivial. To prove necessity, we have

$$
\begin{aligned}
\frac{\beta_w(p,w)}{\beta(p,w)} &= b(w), \\
\frac{\partial}{\partial w}\ln\beta(p,w) &= b(w), \\
\ln\beta(p,w) &= \int b(w)dw+c(p), \\
\beta(p,w) &= e^{\int b(w)dw+c(p)}=\beta_1(p)\beta_2(w),
\end{aligned}
$$

where $\beta_2(w)=e^{\int b(w)dw}$.

The above observations can be summarized as follows:

Proposition 1 *When preferences are quasi-homothetic,*

$$\frac{\partial h}{\partial N}$$

is independent of location if and only if preferences satisfy a separability condition $\beta(p,w)=\beta_1(p)\beta_2(w)$.

Next consider the response of the demand for leisure to wage changes,

$$\frac{\partial h}{\partial w}=a_w(p,w)+b(p,w)+(N+w)b_w(p,w).$$

Again, the goal is to derive conditions under which

$$\frac{\partial h}{\partial w}$$

does not depend on local prices, p. If $b(p,w)=b(w)$, as above, then the only other necessary condition is that $a_w(p,w)$ be independent of p. Now $a_w(p,w)$ is independent of p if and only if it is equal to some function of w only: $a_w(p,w)=f(w)$. Integrating both parts with respect to w, we get $a(p,w)=F(w)+c(p)$. Then the supply of hours of work takes an additively separable form, $h(p,w,N)=c(p)+F(w)+(N+w)b(w)$.

We have established, therefore,

Proposition 2 *When preferences are quasi-homothetic,*

$$\frac{\partial h}{\partial w}\ \textit{and}\ \frac{\partial h}{\partial N}$$

are independent of location if and only if the demand for leisure has the additively separable form

$$h(p,w,N)=c(p)+F(w)+(N+w)b(w). \tag{15}$$

Notice that in equation (15) the effect of local price variation is to simply shift the labor supply function up or down. In this case, it might suffice to merely incorporate location-specific dummies when estimating labor supply functions.[8] Without this separability, however, local price variation would have a fundamental impact on the shape of the labor supply function itself.

[8] In fact, in empirical work on labor supply, researchers generally do not even take this simple step.

These two propositions demonstrate that even in the simple case of quasi-homothetic preferences, rather strong conditions are necessary for location-independent labor supply responses to income and wage changes.

The Stone-Geary example used in the previous section illustrates this point. Indirect utility can be written in the form $V = \alpha(p,w) + (N + w)\beta(p,w)$, where

$$\alpha(p,w) = -\frac{cp\theta A\delta^{\delta}(1-\delta)^{1-\delta}}{p^{\delta}w^{1-\delta}} = -cp^{1-\delta}\theta A\delta^{\delta}(1-\delta)^{1-\delta}\cdot\frac{1}{w^{1-\delta}}, \tag{16}$$

$$\beta(p,w) = \frac{\theta A\delta^{\delta}(1-\delta)^{1-\delta}}{p^{\delta}w^{1-\delta}} = \frac{\theta A\delta^{\delta}(1-\delta)^{1-\delta}}{p^{\delta}}\cdot\frac{1}{w^{1-\delta}}. \tag{17}$$

Since $\beta(p,w)$ is separable in p and w, the separability condition of Proposition 1 is satisfied. Recall from equation (6) that

$$h(p,w,N) = \frac{\delta w - (1-\delta)(N-cp)}{w}.$$

Obviously, this function does not have an additively separable form as required in Proposition 2. So it is not surprising that the derivative of labor supply with respect to nonlabor income, N,

$$\frac{\partial h}{\partial N} = -\frac{(1-\delta)}{w},$$

is independent of p, whereas the derivative of leisure with respect to the wage, w,

$$\frac{\partial h}{\partial w} = \frac{(1-\delta)(N-cp)}{w^2},$$

depends on p.

As noted earlier, labor supply studies generally focus on the responsiveness of labor supply to changes in wages. Here, we want to evaluate how price variations, in addition to changes in wages, affect the results. The ideal experiment would be one in which wages are exogenously shifted in each of many different U.S. cities and in which changes in labor supplied in each city can be traced. Finding data that correspond to such an experiment is a formidable task. The following work instead focuses exclusively on the sensitivity of labor supply to nonlabor income. We can justify this focus with the following result:

Proposition 3 *In general, labor supply, $h(p,w,F)$, depends on the price of the local good, the wage, and full income, $F = w + N$.*[9]

If the key relationship $\frac{\partial h}{\partial w}$ is independent of p, then $\frac{\partial h}{\partial N}$ is independent of p.

To prove this proposition we consider first the effect of a change in nonlabor income on labor supply:

$$\frac{\partial h(p,w,F)}{\partial N} = \frac{\partial h(p,w,F)}{\partial F}\cdot\frac{\partial F}{\partial N} = \frac{\partial h(p,w,F)}{\partial F}.$$

This is independent of price, p, if and only if

$$\frac{\partial h(p,w,F)}{\partial F} = G(w,F). \tag{18}$$

Integrating both sides of equation (18), we then notice that labor supply must have the following additively separable form:

$$h(p,w,F) = g(w,F) + c(p,w) = g(w,w+N) + c(p,w). \tag{19}$$

Similarly, the effect of the change in the wage on labor supply does not depend on p if and only if

$$\frac{\partial h(p,w,F)}{\partial w} = Q(w,F), \tag{20}$$

or, integrating both sides of equation (20),

$$h(p,w,F) = q(w,F) + k(p) = q(w,w+N) + k(p). \tag{21}$$

Compare the additive separability requirements shown in equations (19) and (21). The latter takes the same basic form but is more restrictive. It follows that when

9 Recall that full-time work entails $h = 1$, so that the maximum possible labor income is $w + N$, making full income.

$\frac{\partial h}{\partial w}$ is independent of the local price, p,

$\frac{\partial h}{\partial N}$ is independent of the local price, p.

EMPIRICAL RESULTS

The theoretical considerations outlined in the preceding section suggest that unless preferences are strongly restricted, the responsiveness of labor supply to nonlabor income and to the wage will vary across locations. It is possible, of course, that the differences are insignificant and do not pose a problem for empirical work. We examine this possibility using a dataset of married white women—a group that is likely to have substantial variation in labor supply (e.g., in response to differences in wage, nonlabor income, and possibly local prices). Data used in the analysis are from the 1990 PUMS[10]; data include married non-Hispanic white women, aged 30 to 50, who live in the 50 largest metropolitan statistical areas (MSAs) in the United States.

One goal of this exploration is to see if there are any systematic differences in labor supply related to differences in local prices. We consider the relationship between labor supply and nonlabor income; the latter term is defined as family income minus the woman's own total income. Given previous research on married women's labor supply, an inverse relationship would be expected between nonlabor income and labor supply (i.e., leisure is likely a "normal good.") The question here is whether that relationship differs in a systematic way across cities.

Examining the relationship between nonlabor income and married women's labor supply in cross section is far from "state of the art" in estimating labor supply. Still, it seems a reasonable first pass at the issue, especially given that our focus is not on any estimated relationship per se but on differences in the relationships in expensive and inexpensive urban areas.

In our investigation of the differences in the response of labor supply to the change in nonlabor income, we do not want to specify any parametric form because of concerns that results might be sensitive to the functional form.[11] Instead, we use a nonparametric matching estimator. Two measures of labor supply are used: annual hours of work and an employment participation dummy variable.[12] The data do not allow us to perform this analysis for each city because they do not provide enough support. Instead, we divide the sample roughly into thirds and examine differences between the most "expensive" cities (the 17 MSAs within the top one-third of housing prices) and "inexpensive" cities (the 17 MSAs with the lowest housing prices).

Our comparison of married women's labor supply in inexpensive and expensive cities then follows three additional steps. The first step is to divide households into deciles according to "nonlabor income" (which is predominately the husband's income). Then within each decile we compare the labor supply of women who live in the expensive cities relative to the labor supply of women who live in inexpensive cities. The goal is to compare the labor supply of otherwise similar women, so we use an estimator that matches women with exactly the same age and level of education. Separate analyses also are conducted for women with high school education and college education. Thus, the second step is to match women living in an expensive city with corresponding women living in inexpensive cities (i.e., we match women in each nonlabor income decile, d_i ($i = 1,\ldots,10$), with age and education vector $x = X$, to women with these same characteristics living in inexpensive cities). In the analysis that centers on annual work hours, this is

$$(22)\quad \Delta(X,d_i) = E(h_1 \mid x = X,d_i) - E(h_0 \mid x = X,d_i),$$

where h_1, h_0 are annual hours of work in expensive and inexpensive cities, respectively. In the absence of selection, this might be taken to be the causal effect on labor supply (measured in hours per year) of living in an expensive city relative to

[10] Data were provided by the Minnesota Population Center (Ruggles et al., 2008).

[11] See, for example, DaVanzo, DeTray, and Greenberg (1973).

[12] We also repeated the analysis with several other measures of labor force participation, such as an indicator of full-time employment. The results remain essentially the same.

Table 1

Differences in Annual Hours and Participation Rates Between Expensive and Inexpensive Locations by Nonlabor Income Deciles

	All women		Women with a high school diploma		Women with a college degree	
Nonlabor income decile	Change in annual hours	Change in participation rates	Change in annual hours	Change in participation rates	Change in annual hours	Change in participation rates
1	–117.34	–0.04	–136.1	–0.04	–78.08	–0.02
	(14.23)	(0.0065)	(24.57)	(0.012)	(34.88)	(0.016)
2	–75.46	–0.01	–75.72	0.00	–99.43	–0.02
	(14.32)	(0.0063)	(24.36)	(0.011)	(36.47)	(0.016)
3	–54.14	–0.01	–19.42	0.00	–46.71	–0.01
	(13.74)	(0.0060)	(23.39)	(0.012)	(33.98)	(0.015)
4	–15.14	0.00	–28.97	–0.01	–20.59	0.00
	(13.88)	(0.0062)	(23.63)	(0.012)	(37.16)	(0.016)
5	–20.68	0.01	–51.79	0.00	–13.31	0.03
	(13.31)	(0.0063)	(24.14)	(0.012)	(34.57)	(0.015)
6	2.59	0.02	–39.52	0.00	59.98	0.05
	(13.66)	(0.0068)	(24.14)	(0.013)	(31.66)	(0.015)
7	12.47	0.01	–16.11	0.00	85.6	0.03
	(14.38)	(0.0072)	(24.79)	(0.013)	(30.99)	(0.015)
8	83.55	0.05	81.95	0.05	139.38	0.08
	(14.62)	(0.0076)	(26.78)	(0.014)	(30.24)	(0.015)
9	83.61	0.04	88.98	0.03	128.59	0.06
	(15.80)	(0.0083)	(33.44)	(0.017)	(30.84)	(0.016)
10	82.59	0.04	15.74	0.00	172.35	0.07
	(18.45)	(0.0098)	(41.52)	(0.023)	(28.04)	(0.015)

NOTE: Authors' calculations, based on 5 percent 1990 PUMS data. The sample consists of white, non-Hispanic married women, aged 30 to 50. Bootstrapped standard errors using 999 replications are reported in parentheses.

an inexpensive city. The third step is to average the quantity in equation (22) across all women in each decile d_i:

$$\Delta_n(d_i) = \int \Delta(x \mid d_i) dF_n(x \mid i), \tag{23}$$

where $dF_n(x \mid d_i)$ is the national distribution of x in the decile d_i.

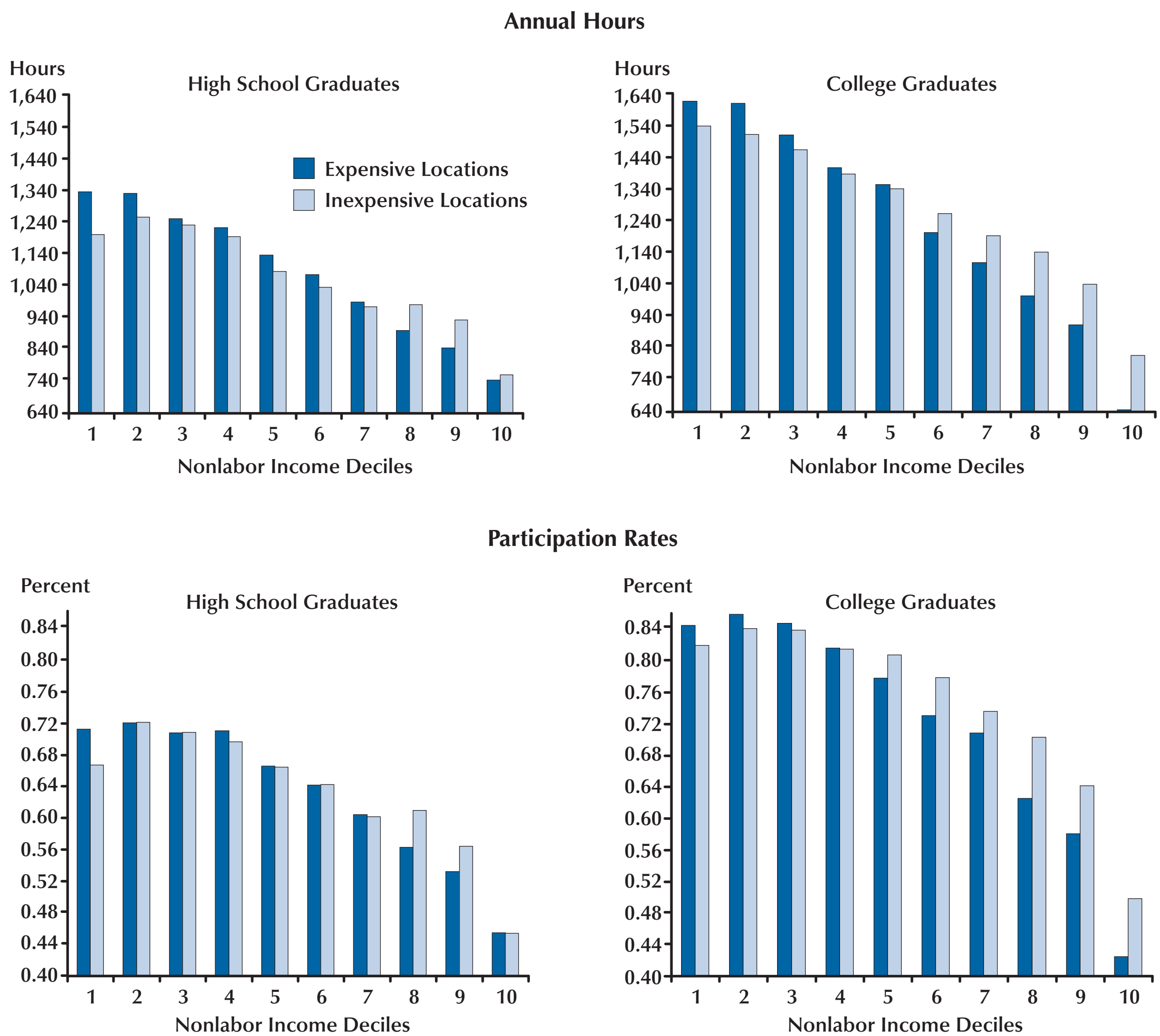
The analysis is repeated using a second measure of labor supply—a labor force participation dummy variable. When these empirical exercises are performed separately for women with a high school diploma and those with a college degree, x, is simply an age vector.

Results are reported in Table 1. The difference in annual hours of work between women living in expensive and inexpensive cities is substantial (and statistically significant) for many of the nonlabor income deciles. For example, ninth-decile women in expensive cities work considerably longer hours than corresponding women in inexpensive cities. College-educated women in this decile average 129 more work hours, whereas women with a high school education work an average of 89 hours more.

An apparent and striking pattern is shown in Table 1 and Figure 1. First, as might be expected, among these married women, leisure appears to be a normal good; women with higher levels of outside income generally work fewer hours per year and have lower labor force participation rates.

Figure 1

Variation Between Expensive and Inexpensive Locations in Annual Hours and Participation Rates, by Nonlabor Income Decile



More important, for our purposes, is that the relationship between nonlabor income and labor supply is quite different for expensive and inexpensive cities. At the very lowest levels of nonlabor income (e.g., deciles 1 and 2), women in expensive cities have lower labor supply than women in inexpensive cities. The opposite is essentially true for women in the high nonlabor income deciles; among women with high nonlabor income, labor force participation and average hours worked are higher in expensive cities than in inexpensive cities.

In short, the labor/leisure choice appears to not conform to the additively separable form described in Proposition 2; local prices do not merely shift labor supply up or down. The derivative

$$\frac{\partial h}{\partial N}$$

is generally negative (at least beyond the lowest decile levels of N) and is smaller (in absolute value) in the expensive city. This generalization holds true for both high school– and college-educated women.

Also, as noted, results are similar when "average hours" or "labor force participation rates" are used as the measure of labor supply. Of note, in these cities 66 percent of high school–educated women and 70 percent of college-educated women are employed on average. Thus, differences of 5 to 7 percentage points between expensive and inexpensive cities represent differentials of 8 to 10 percent, which seem (to us) quite substantial.

Our nonparametric approach does have one disadvantage: The nonlabor income distribution within each decile might differ somewhat for women in expensive cities. An alternative flexible parametric approach to estimation, described in the Appendix, provides nearly identical inferences.

Our empirical findings are roughly consistent with theoretical predictions in Case 2. In that equilibrium example with Stone-Geary preferences, the responsiveness of labor supply to nonlabor income must be greater in inexpensive (low-productivity) cities than expensive (high-productivity) cities.

CONCLUSION

We describe a simple model to demonstrate that the effects of wage and nonlabor income on labor supply typically differ by location. In particular, we show the derivative of the labor supply with respect to nonlabor income is independent of price only when labor supply takes a form based on an implausible separability condition.

Empirical evidence demonstrates that the effect of price on labor supply is not a simple "up-or-down shift" that would be required to meet the separability condition in our key proposition. For example, among women with low nonlabor income, living in an inexpensive city is associated with *higher* labor force participation and *longer* work hours, whereas among women with high nonlabor income, living in an inexpensive city is associated with *lower* labor force participation and *shorter* work hours.

This work has a number of implications for empirical strategies in estimating labor supply and other policy research. First, our research makes clear that empirical work should never use cross-sectional variation in wages to estimate parameters in labor supply models. We document significant differences for married women in quantity of labor supplied across cities that may have little connection with behavioral responses to cross-sectional variation in wages.

Second, because labor supply elasticities vary by location, researchers must be careful in interpreting results based on instrumental variable (IV) strategies. For example, suppose an IV approach is used in which the IV is the price of coal. Variation in the price of coal arguably serves as an excellent source of wage variation in the coal industry, but the resulting estimates of the effect on labor supply would apply only for regions where the coal industry is a major employer. If local prices differ in those regions from other parts of the country, the estimated relationships will not be generalizable to the entire country.

Third, using a back-of-the-envelope example, we show that the evidence in Table 1 is consistent with the possibility that wage elasticities or labor supply (for married women) are quite different across cities. Notice that the Slutsky equation, in elasticity form, gives the relationship

$$\varepsilon_w = \varepsilon_w^H + \left[\frac{wh}{N}\right]\varepsilon_N, \qquad (24)$$

where ε_w is the observed wage elasticity of supply, ε_w^H is the corresponding Hicksian elasticity (reflecting the pure substitution effect), and ε_N is the elasticity of labor supply with respect to nonlabor income. Now consider college-educated married women at the median level of nonlabor income. If we take as causal the relationship drawn in Figure 1, moving from the fourth to sixth deciles in income we would estimate a nonlabor income elasticity, ε_N, of –0.46 in the expensive cities and –0.29 in the inexpensive cities. Suppose that the Hicksian elasticity, ε_w^H, is 0.50 (and is the same in both cities). We estimate that for the average woman at the fourth decile wh/N is 0.57 in inexpensive cities and 0.61 in expensive cities.[13] Thus, the uncompensated labor supply elasticity is more than a third higher in expensive cities than inexpensive cities, 0.33 versus 0.24.

Fourth, as an example of an application to policy-related research, locational differences may occur in the response of female labor supply to changes in taxes. Changes in income taxes, for instance, would have different effects in different cities. A closely related implication centers on the analysis of social welfare policy. (Recall, for example, that wives of husbands with *low* earnings work less in more expensive cities.) We believe that further analysis of policy implications is warranted.

REFERENCES

Abbott, Michael and Ashenfelter, Orley. "Labour Supply, Commodity Demand and the Allocation of Time." *Review of Economic Studies*, 1976, *43*(3), pp. 389-411.

Ashenfelter, Orley and Ham, John C. "Education, Unemployment, and Earnings." *Journal of Political Economy*, 1979, *87*(5), pp. 99-116.

Blundell, Richard and MaCurdy, Thomas. "Labor Supply: A Review of Alternative Approaches," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*. Princeton, NJ: Princeton University Press, 1999, Vol. 3, pp. 1559-95.

Chen, Yong and Rosenthal, Stuart S. "Local Amenities and Life-Cycle Migration: Do People Move for Jobs or Fun?" *Journal of Urban Economics*, November 2008, *64*(3), pp. 519-37.

DaVanzo, Julie; DeTray, Dennis N. and Greenberg, David H. "Estimating Labor Supply Response: A Sensitivity Analysis," publication No. R-1372-OEO. Santa Monica, CA: The RAND Corporation, 1973.

Gabriel, Stuart A. and Rosenthal, Stuart S. "Quality of the Business Environment Versus Quality of Life: Do Firms and Households Like the Same Cities?" *Review of Economics and Statistics*, 2004, *86*(1), pp. 438-44.

Haurin, Donald R. "The Regional Distribution of Population, Migration, and Climate." *Quarterly Journal of Economics*, 1980, *95*(2), pp. 293-308.

Killingsworth, Mark R. and Heckman, James J. "Female Labor Supply: A Survey," in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*. Princeton, NJ: Princeton University Press, 1986, Vol. 1, pp. 103-204.

Masters, Stanley H. and Garfinkel, Irwin. *Estimating the Labor Supply Effects of Income Maintenance Alternatives*. New York: Academic Press, 1977.

Pencavel, John. "Labor Supply of Men: A Survey," in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*. Princeton, NJ: Princeton University Press, 1986, Vol. 1, pp. 3-102.

Roback, Jennifer. "Wages, Rents, and the Quality of Life." *Journal of Political Economy*, 1982, *90*(6), pp. 1257-78.

Ruggles, Steve; Sobek, Matthew; Alexander, Trent; Fitch, Catherine; Goeken, Ronald; Hall, Patricia; King, Miriam and Ronnander, Chad. *Integrated Public Use Microdata Series: Version 4.0* (machine-readable database). Minneapolis, MN: Minnesota Population Center, 2008; usa.ipums.org/usa/.

[13] In fact, the ratio of women's earnings to nonlabor household income (primarily men's earnings) is larger in expensive cities than in inexpensive cities at every decile.

APPENDIX

The empirical inferences in Table 1 are based on an entirely nonparametric approach. We divided our sample into 10 nonlabor income deciles and compared labor supply across women within each of these cells. Our primary finding is that for women in low nonlabor income deciles, the labor supply is *lower* in expensive cities than in inexpensive cites, whereas for women in high nonlabor income deciles, labor supply is *higher* in expensive cities than in inexpensive cities.

Here we present a flexible parametric approach that leads to this same inference. We estimate labor supply regressions with the independent variables age (entered as 21 dummy variables for each age, 30 to 50 years inclusive) and nonlabor income (entered as a fourth-order polynomial). We estimate regressions—separately for high school–educated women and college-educated women, as well as for each labor supply variable (employment and hours worked)—using the sample of women from the expensive cities. We similarly estimate corresponding regressions for the sample of women from the inexpensive cities. Then for each woman i who lives in the expensive cities, we estimate the outcome of interest $\hat{y}_{1i}$ (e.g., "predicted" employment, or "predicted" hours worked) using the regression parameter from the *expensive* city, and similarly estimate $\hat{y}_{0i}$ using regression parameters from the *inexpensive* city. Finally, we form the estimated gap,

$$\hat{\Delta}_i = \hat{y}_{1i} - \hat{y}_{0i}\,,$$

for each individual. Notice that this last quantity is the "impact of the treatment on the treated," where the "treatment" is location in an expensive city rather than an inexpensive city.

To summarize findings in a manner comparable to Table 1, we aggregate estimates into deciles of nonlabor income. Results are presented in Table A1. Bootstrapped standard errors using 999 replications are reported in parentheses.[14]

[14] Bootstrap procedure in this case involves 999 replications of generating a random sample with replacement from the original dataset and estimating the parameter of interest for that sample. After 999 replications, we have a sampling distribution of the parameter estimate. The standard deviation of that distribution is the standard error of the parameter estimate.

Table A1

Differences in Annual Hours and Participation Rates Between Expensive and Inexpensive Locations by Nonlabor Income Deciles, Parametric Approach

	Women with a high school diploma		Women with a college degree	
Nonlabor income decile	Change in annual hours	Change in participation rates	Change in annual hours	Change in participation rates
1	–128.7	–0.034	–118.1	–0.027
	(22.04)	(0.0110)	(34.23)	(0.0143)
2	–93.4	–0.021	–72.5	–0.016
	(12.42)	(0.0066)	(17.76)	(0.0079)
3	–68.6	–0.013	–36.6	–0.002
	(11.10)	(0.0059)	(16.07)	(0.0074)
4	–47.1	–0.005	–9.5	0.009
	(10.82)	(0.0056)	(15.23)	(0.0071)
5	–28.1	0.001	19.1	0.021
	(10.26)	(0.0056)	(14.59)	(0.0066)
6	–2.1	0.01	46.5	0.032
	(11.15)	(0.0056)	(14.18)	(0.0066)
7	23.8	0.019	76.5	0.045
	(12.73)	(0.0061)	(14.59)	(0.0071)
8	55.3	0.030	108.6	0.058
	(15.28)	(0.0077)	(17.27)	(0.0082)
9	87.5	0.042	143.5	0.075
	(20.48)	(0.0102)	(20.89)	(0.0099)
10	81.6	0.036	123.1	0.066
	(38.06)	(0.0207)	(30.26)	(0.0151)

NOTE: Authors' calculations, based on 1990 PUMS data. The sample consists of all married, white, non-Hispanic women between the ages of 30 and 50 inclusive. The covariates are nonlabor income and age. Using a fourth-order polynomial, we use the sample of women from expensive cities to estimate the outcome of interest, which we denote $\hat{y}_{1i}$ for the ith women. Using the sample of women from inexpensive cities, we estimate parameters for a fourth-order polynomial and then evaluate the function using the covariates of women from the expensive city sample, which we denote $\hat{y}_{0i}$ for the ith women. We then form the parameter for the "impact of treatment on the treated" as $\hat{\Delta}_i = \hat{y}_{1i} - \hat{y}_{0i}$. We then aggregate estimates into deciles of nonlabor income. Bootstrapped standard errors using 999 replications are reported in parentheses.

What Happened to the U.S. Stock Market? Accounting for the Past 50 Years

Michele Boldrin and Adrian Peralta-Alva

The extreme volatility of stock market values has been the subject of a large body of literature. Previous research focused on the short run because of a widespread belief that in the long run the market reverts to well-established fundamentals. The authors' research suggests this belief should be questioned. First, they show actual dividends cannot account for the secular trends of stock market values. They then consider a more comprehensive measure of capital income, which displays large secular fluctuations that roughly coincide with changes in stock market trends. Under perfect foresight, however, this measure fails to properly account for stock market movements. The authors thus abandon the perfect foresight assumption and instead assume that forecasts of future capital income are performed using a distributed lag equation and information available up to the forecasting period only. They find that standard asset-pricing theory can be reconciled with the secular trends in the stock market. This study, nevertheless, leaves open an important puzzle for asset-pricing theory: The market value of U.S. corporations was much lower than the replacement cost of corporate tangible assets from the mid-1970s to the mid-1980s. (JEL E25, G12)

Federal Reserve Bank of St. Louis *Review*, November/December 2009, *91*(6), pp. 627-46.

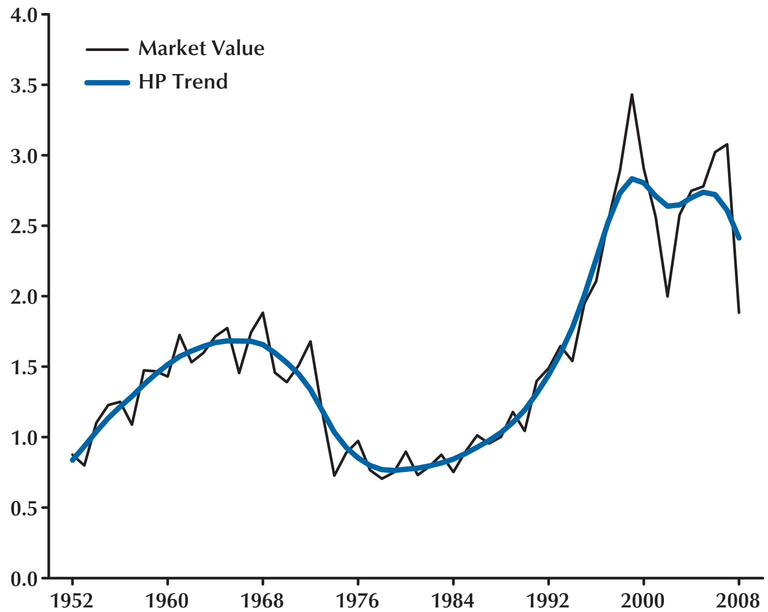
Standard (consumption-based) asset-pricing models have a hard time explaining *high-frequency* fluctuations in stock market values, given observed fluctuations in fundamentals. The anomalies these models face have been labeled in a variety of ways—all ending with the word "puzzle." Various solutions have been suggested, but it seems no more than a handful of researchers have deemed any of these satisfactory. Campbell (2003) provides a summary of the various puzzles and solutions proposed by the consumption-based asset-pricing literature, and Boldrin, Christiano, and Fisher (2001) offer the solution that, so far, we find the least unconvincing. While the search continues, it becomes more apparent that the hope of capturing the stock market's short-run gyrations by appropriately filtering the quarterly movements of aggregate consumption is unlikely to be realized. The question then is this: If consumption-based models cannot do the job, what can? In this article we contribute our two cents to the collective effort of answering this most difficult question.

Before completely abandoning the standard model, we ask a seemingly less challenging but more fundamental question. Namely, can the standard (that is to say, net present-value–based) economic theory of asset pricing account for the very large *low-frequency* fluctuations in the aggregate stock market valuation of U.S. corporations? In other words, if we ignore short-term movements and look only at very long-run trends—those persistent enough to last at least five, and generally

Michele Boldrin is the Joseph G. Hoyt Distinguished University Professor in the Department of Economics at Washington University in St. Louis and a research fellow at the Federal Reserve Bank of St. Louis. Adrian Peralta-Alva is an economist at the Federal Reserve Bank of St. Louis.
Chanont Banternghansa provided research assistance.

Figure 1

Market Value of U.S. Corporations (Relative to Corporate GDP)



more, years—is the standard model capable of correctly explaining/predicting those trends? As far as we know, this question has seldom, if ever, been addressed systematically. Yet, it is relevant for an assessment of asset-pricing models for this reason: There is a widespread belief that, although the standard model may miss a few short-term bumps, in the long run the market will revert to well-established fundamentals. Our research suggests this belief should be questioned.

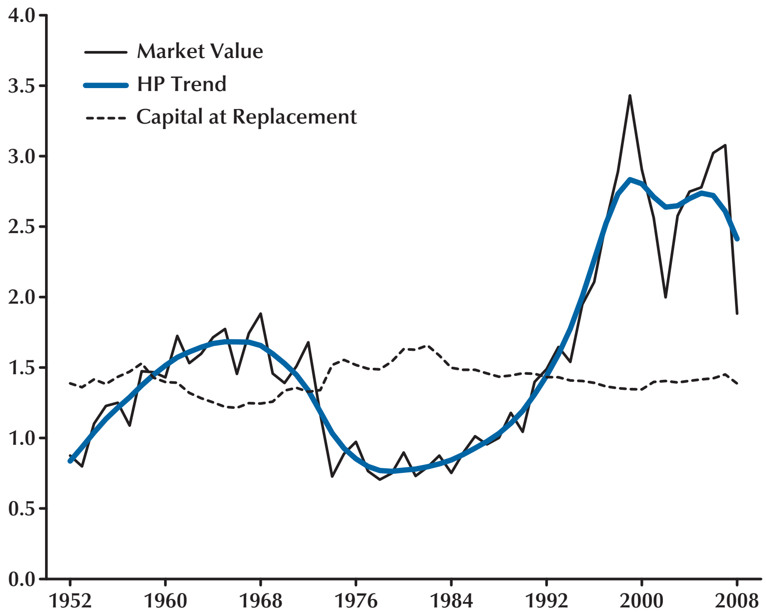
Our analysis is structured as follows. First, we document the secular trends in the value of U.S. corporations. Available data rule out the possibility that fluctuations in market value might have been caused by fluctuations in corporate assets. Then, we study the implications of a fundamental asset-pricing equation according to which asset prices equal the expected discounted present value of returns. As is common in the literature, to test the implications of the theory, we use aggregate data (either from publicly traded firms or from the overall corporate sector). First, we show that, under perfect foresight, the standard approach of computing the present value of actual stock market dividends or returns accounts for little of the secular trends of the U.S. stock market. Since dividend payments may respond to complicated corporate finance considerations, we then study whether movements in the whole of shareholders' income may better account for stock market trends. We find that this is not the case. Finally, we drop the perfect foresight assumption and assume that shareholders make forecasts based only on available information and a distributed lag equation. As we show, this assumption together with the fundamental asset-pricing equation explains much of the secular trends of the U.S. stock market.

THE SECULAR TRENDS OF THE VALUE OF CORPORATE CAPITAL

Figure 1 summarizes key features of the data. Our data appendix details the sources and methods used to construct this and all other figures in this paper. To normalize for economic growth, we

Figure 2

Market Value and Replacement Value of Corporate Capital (Relative to Corporate GDP)



focus on the ratio of the market value of corporations to corporate value added (or other measures of aggregate output, when appropriate); we refer to this as the "market ratio." The black line in Figure 1 shows the market ratio during the past 55 years (based on annual data). The blue line shows the low-frequency movements in the market ratio by means of the trend generated by the Hodrick-Prescott (HP) filter. The use of the HP filter is standard in the literature and it is applied to all variables in this paper. Almost undistinguishable patterns result when other reasonable long-run filters are used.

As the figure shows, after two decades of growth the market ratio declined by 50 percent during 1973-74 and stagnated until the mid-1980s. From 1985 to 2000 it more than tripled, only to collapse again by 2001. Since then, the market ratio has fluctuated around 2.7, taking a gigantic drop (only partially reported in the figure and now partially recovered) during recent months. These are extremely large movements by any metric and dwarf the also substantial oscillations at the quarterly to yearly frequencies. The question is this: What kind of economic rationale drives such impressive swings?

The most elementary model we can think of analyzes aggregate production and capital accumulation over time. This type of model considers an aggregate firm producing national consumption (in fact, gross national product [GNP]) by employing capital, K, and labor, L, under a constant returns-to-scale production function, $F(K,L)$. The resource and wealth constraints for this economy are

$$F(K_t, L_t) = C_t + I_t;\ K_{t+1} = (1-\mu)K_t + I_t.$$

In this environment, consumption and investment (and therefore capital) are interchangeable on a one-to-one basis. Hence, the price (or value) of the capital stock, measured in units of the consumption good, is always one. It follows that the market ratio must equal the physical capital/output ratio implied by the aggregate production function $F(K,L)$. This most elementary explanation is immediately ruled out by the data. In Figure 2, we add a third line to Figure 1, which shows the ratio of the replacement value of corporate capital to corporate gross domestic product (GDP) (that is to say, K_t/Y_t). It shows a remarkable stability compared with the market ratio: Although some long-run oscillations are visible, they are of about one order of magnitude smaller than those of the market ratio and in the opposite direction. In summary, an explanation for the huge swings in the market ratio needs to be found somewhere beyond the actual stock of capital owned by U.S. corporations or the cost of producing it.[1]

PERFECT FORESIGHT OF FUTURE DIVIDENDS

If oscillations in the market value of capital cannot be explained in terms of either its cost or its quantity (relative to labor and/or output), maybe they can be explained in terms of "value." The market ratio increases (decreases) because the capital stock used by corporations becomes less (more) productive, hence yielding more (less) profits to its owners. According to this principle, the market value of corporate capital is determined by looking forward and not backward: Independent of both how capital intensive production processes may be and the cost of installed machines, the market value will increase if capital is productive and its owners expect it to yield profit; and it will decrease in the opposite case. In summary: Standard asset-pricing theory says that the market ratio is a forecast. The questions are (i) the market ratio is a forecast of what and, (ii) how correct this forecast has turned out to be? We will concern ourselves with these two questions for the rest of this paper.

To begin answering them, we establish the simplest possible framework of analysis in which the value of corporations is equal to the value of what their capital will produce and earn. We contemplate dynamic stochastic economies that are, on a period-by-period basis, subject to a vector of exogenously given shocks. Such shocks—which may include changes in productivity, demand, taxes, and other factors—are the sources of uncertainty that the forward-looking agents must examine to price assets on the basis of their expected future returns. Let E_t be the expectation operator, taken with respect to the probability distribution of future shocks on the basis of the information available at time $t = 0,1,2,\ldots$; d_t are the dividends paid by the firm to shareholders, τ_t^d and τ_t^V the dividends and capital gains income tax rates, and V_t the market value of the firm, all as of period t. Finally, let p_{t+i} be the stochastic discount factor of future consumption—that is, the value today of one unit of consumption obtained, in some state of the world, during the future period $t+i$, $i = 1,2,3,\ldots$ The following relation holds:

$$(1)\quad V_t = \left(1-\tau_t^d\right)d_t + \mathrm{E}_t\left[\sum_{i=1}^{T}\left(1-\tau_{t+i}^d\right)p_{t+i}d_{t+i}\right] + \mathrm{E}_t\left[\left(1-\tau_{t+T}^V\right)p_{t+T}V_{t+T}\right],$$

Here, T denotes any arbitrary positive number (including plus infinity). This formula states that the market value of a firm should equal the expected present discounted value of the future stream of (after-tax) shareholders' income it generates plus the (after-tax) capital gains/losses that result from selling the share at some future period.

It is important to note that the fundamental asset-pricing equation (1) holds in a wide range of economic models. Indeed, different branches of the literature have emerged from varying the key assumptions and methods for deriving predictions from this equation. The consumption-

[1] A slightly more sophisticated version of this model, which allows for changes in the relative price of capital (q), can be evaluated but yields similar (counterfactual) results. Suppose the investment good is produced with a different technology from the one producing the consumption good (see, e.g., Boldrin, Christiano, and Fischer, 2001). In this case, the resource and wealth constraints read as

$$F\left(K_t^c, L_t^c\right) = C_t;\ G\left(K_t^k, L_t^k\right) = I_t;$$
$$Y_t = C_t + q_t I_t;\ K_{t+1} = (1-\mu)K_t + I_t;$$

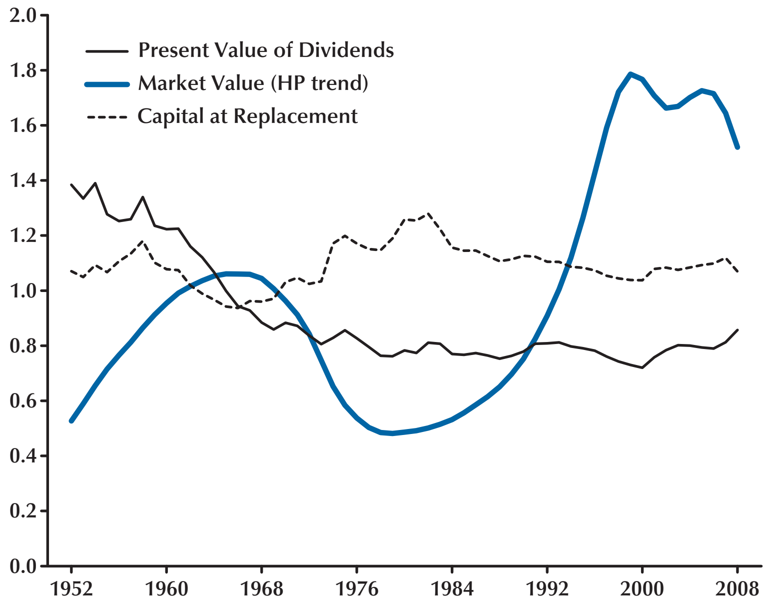
and the market ratio equals

$$\frac{q_t K_t}{F\left(K_t^c, L_t^c\right) + q_t G\left(K_t^k, L_t^k\right)}.$$

Hence, either "biased" technological change or changes in sectoral factor intensities could bring about a change in the relative price of capital. Moreover, the market ratio may move because the capital intensity of aggregate production moves or because the relative price of capital oscillates. Notice, however, that the market ratio predicted by the model, expression (2), should still correspond to the capital output ratio of the United States (just as the simpler one-sector model).

Figure 3

Present Value of Dividends versus Market and Replacement Values (Relative to Corporate GDP)



based asset-pricing literature, for instance, assumes dividends and consumption are exogenously given processes. In this literature, the interaction of the stochastic discount factor and the dividend process are the key forces driving the volatility of asset prices. The production-based asset-pricing literature, in turn, develops the asset-pricing implications of models where consumption and dividends are endogenously determined. Finally, the present-value pricing literature considers long holding periods for shares (high values for T, in our notation) and explores the asset-pricing implications of alternative measurements for dividends and long-run discount factors.

To derive the exact quantitative implications of the asset-pricing equation (1), one would need to measure all of the possible time series of taxes, discount factors, capital income, and, in particular, the probabilities the market assigns to all possible future states of the world. Determining the latter directly, at any given point in time, is an impossible task because the theory, per se, admits the most arbitrary set of expectations for the participating agents. A common benchmark followed in the literature (e.g., Shiller, 2005) consists of assuming a constant discount factor and approximating capital income by actual dividends. Crucially, perfect foresight on dividends is commonly assumed as well: The market prices are supposed to be based on exact forecasts of the realized dividends; hence, realized dividends can be used in the computations. Because these are open-ended models, existing analyses typically complement the perfect foresight hypothesis with the additional assumption that dividends will grow at some average rate for the infinite future. Generally, this literature also abstracts away from fluctuations in the tax on dividend income.

This approach does not go very far in accounting for equity price movements. A representative illustration of the predictions from the theory under the aforementioned assumptions, based on the dividend data compiled by Shiller (2005), is presented in Figure 3. For comparison, the ratios displayed below have been normalized so that their 1960-72 average is equal to one. (We will follow this normalization procedure throughout this paper.) These computations assume a 7 percent discount rate and a terminal growth rate for dividends of 3 percent for the infinite future following period *t*. We use a constant discount rate to simplify our presentation, but our results do not change much if we use instead a discount rate based on a power function of consumption growth, as is standard in consumption-based asset-pricing theory. Our assumption of 3 percent terminal growth for dividends is also incorrect since the dividend-to-GDP ratio has been decreasing over time, while GDP has been growing at an average of about 3 percent for most of the period. The perfect foresight assumption implies that the market ratio should have been very high earlier on—as dividends were a high percentage of corporate GDP in the 1950s and discounting matters—to then subsequently decrease and remain quite stable (as dividends' growth rates stabilized from the mid-1970s onward). This makes the large oscillations that took place between 1970 and today impossible to explain on the basis of dividend payoffs, perfect foresight, and a stable dividends-to-GDP ratio. Our first conclusion is that one, or more, of these three assumptions—dividends are the payoff to be forecasted, they are forecasted correctly, and they grow at roughly the same rate as corporate GDP—must be disposed of.

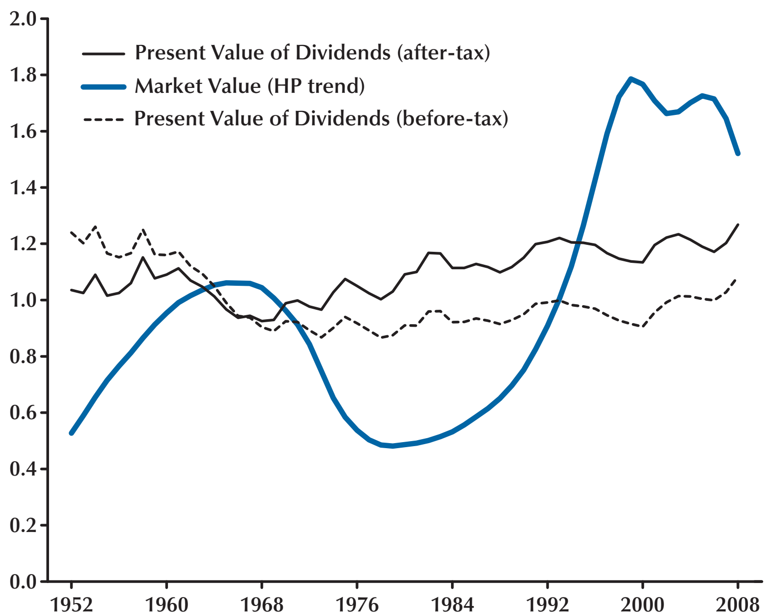
The fact that equity prices increased so rapidly during the late 1990s and that the value of dividends did not has been interpreted by some as evidence of irrational exuberance. This is not necessarily correct: There may have been "exuberance," but it need not have been "irrational" insofar as the information available to the market did not need to be of sufficient quality or amount to correctly forecast future dividends. We will return to this point later, as the issue of what the market can and cannot "forecast correctly" is at the root of the problem we are addressing. In any case, an "irrational bubble" might have been partially behind equity prices during the mid-1990s, but notice that even after both the 2000 and recent stock market crashes, the market ratio is much higher than during the 1980s: So, what is it that the market ratio is therefore "forecasting"? Similarly, the issue of why equity prices were so low in the early 1950s and in the mid-1970s and -1980s is not frequently addressed in the financial literature either, which again brings up a similar question: What was the market ratio "forecasting" during those periods? Not dividends—or, at least, not correctly—because the net present value of actual future dividends is above the market ratio between the early 1970s and the early 1990s.

As we will argue later, the low market valuations in the middle period appear to be the hardest to understand. Notice in passing that this hypothetical market ratio, computed only on the basis of observed dividends, is much closer to the replacement value of capital than the actual market ratio. In other words, if the stock market had really valued corporations on the basis of actually realized dividends, the market and replacement values of corporate capital would have been relatively close during the 56 years we study, and the only long-run puzzle would be a persistent difference, in levels, between the replacement cost of capital and the present value of the dividends it has been generating. Such a puzzle could be easily solved, though, by lowering the discount rate below the 7 percent value we used in the reported calculation. But, apparently, this is not what the stock market did.

We now refine this "perfect foresight" methodology by modifying the object supposedly forecasted by the market ratio. First, McGrattan and Prescott (2005) document important changes in the taxation of dividend income and investment subsidies that may account for some of the observed fluctuations in equity values. We therefore recompute the implications of the theory by adjusting for the varying rate of dividends' taxation. The results are shown in Figure 4, and they are not good. Which leads us to repeat the observation made earlier: The perfect foresight assumption, when applied to the valuation of future payment streams, imposes strong restrictions on the model's predictions. In particular, by eliminating any learning process and assuming the market "knows" future events much earlier than they materialize, it tends to "front-load" all historical changes, producing (thanks to discounting) very flat predictions. In summary, if agents can more or less perfectly forecast all relevant vari-

Figure 4

Present Value of Dividends (Before- and After-Tax) versus Market Value (Relative to Corporate GDP)



ables, the long-run swings of the market ratio make no sense whatsoever. This suggests that the problem may not be with "what" the market forecasts but with "how" it does it.

It may be possible to offer arguments in favor of the perfect foresight assumption for economic fundamentals (such as dividends and discount factors). Assuming perfect foresight in policy variables such as taxes seems much harder to do. Indeed, McGrattan and Prescott's (2005) analysis studies the impact of an unexpected and permanent change in taxes.

We thus recompute the predictions for the theory under the assumption of perfect foresight on dividends and interest rates, but assuming that, at each period, a new tax rate arrives unexpectedly and this rate is believed to persist into the infinite future. As first noted by Bian (2007), such changes in dividend income taxes can only (very) partially account for the higher values of the market ratio from 1994 to 2008. However, in this case, the size of the increase is too small and its timing is way off. Figure 5 suggests that the stock market undervalued corporations between 1952 and 1961 and then, again, between 1970 and 1992, while some kind of "exuberance" (irrational or not, we will see) has driven the market ratio from about 1996 to the present. Simply, even after allowing for large tax surprises, the net present value of future dividends provides a very poor explanation of what happened to the market ratio.

Symmetry would require assuming that a new (permanent) growth rate for dividends also arrives, unexpectedly, at each period. Under these conditions—namely, a random walk growth rate for dividends, g, and a constant discount factor, r—the asset-pricing equation will simply predict that the market ratio moves proportionally with the dividend growth rate. In fact, the random walk hypothesis implies that the market value should equal today's (after-tax) dividends divided by $r_t - g_t$. Figure 6 shows the growth rate of real dividends and averages (of different lengths) of past growth rates. Notice, first, that the growth rate of

Figure 5

Present Value of Dividends (Before and After Unexpected Tax Changes) versus Market Value (Relative to Corporate GDP)

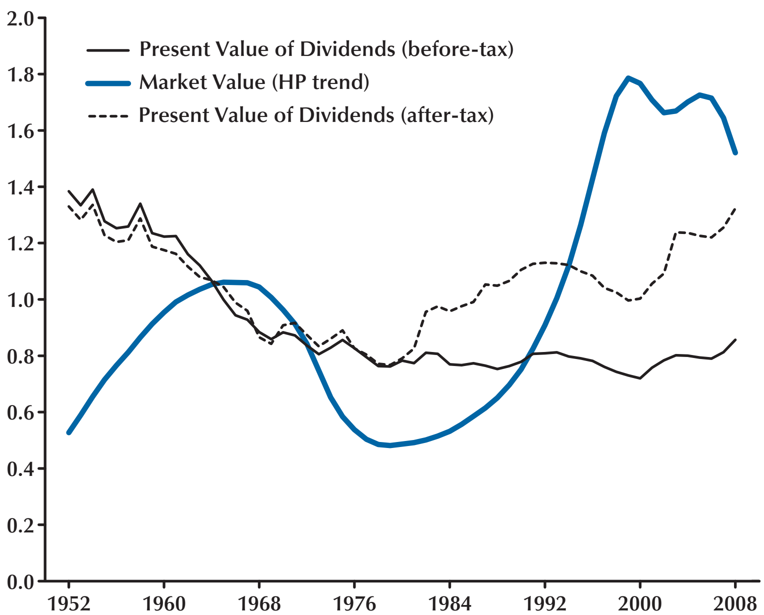
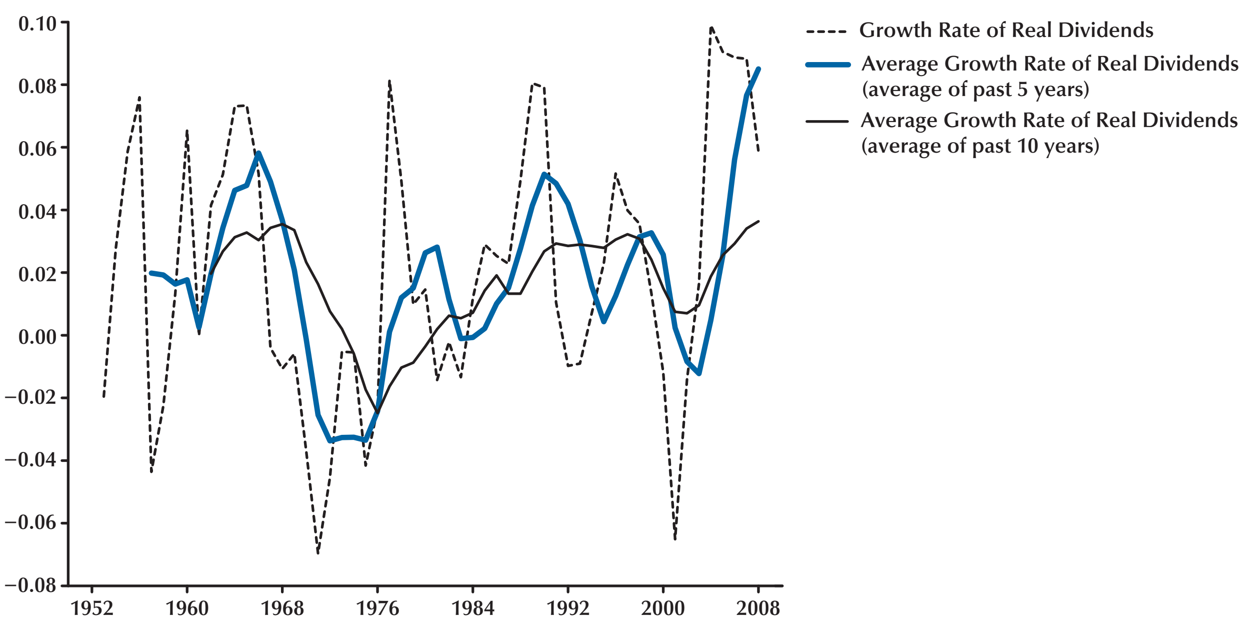


Figure 6

Growth Rate of Dividends (Actual and Averages of Past Years)



dividends is fairly volatile. Second, up to the mid-1980s, changes in the dividend growth rate roughly coincide with changes in the trend of the market ratio. The growth rate of dividends goes down around 1968, and so does the market ratio. Similarly, dividend growth is low through the mid-1970s and it does not recover until the mid-1980s. The market ratio follows similar patterns. Dividend growth does not have any specific trend, on average, during 1992-2008, but it displays higher volatility. While the big increases in the market ratio of the mid-1990s and later are hard to be accounted for by trends in dividend growth, it is surprising how well the 5- and 10-year averages of the dividends' growth rate mimic the long-run gyrations of the market ratio.

It is important to stress that, although changes in the (average trend of the) dividend growth rate are positively related to changes in the trend of the market ratio, the growth rates of actual dividends are very often negative. Of course, assuming that dividends will grow at a negative rate for the infinite future is not very realistic, which makes us return to our fundamental question. If it is not a forecast of actual dividends paid, then the market ratio forecast is a forecast of what? The sections that follow refine the production-based asset-pricing model to provide one possible answer.

PERFECT FORESIGHT OF TOTAL CAPITAL INCOME

Actual dividends paid, in light of our previous analysis, cannot help explain any of the big historical swings in the market ratio. It is not clear, however, that one should use actual dividends in testing the theory. In particular, actual dividend payments may respond to additional considerations such as informational asymmetries, principal-agent revelation mechanisms, fiscal incentives other than those captured by the taxation of dividends and capital gains, and so on. (Easterbrook, 1984, and Feldstein and Green, 1983, review some of the relevant literature.) In trying to determine how far the fundamental asset-pricing equation (1) can take us, it seems more appropriate to abstract from dividend payment considerations and start instead from a simple framework whereby firms' net worth equals the present value of *all* shareholders' income.

The earlier model of production can be adapted to this end by assuming that the aggregate firm chooses capital, labor, and investment to maximize the net present value of shareholders' income. In this new version of the model, shareholders' income is endogenously determined. Further, shareholders are the residual claimants of corporate value added after compensation of employees, corporate income taxes, and gross investment are taken care of. It follows that income accruing to shareholders in period t is

$$d_t = Y_t - w_t L_t - I_t - \text{taxes},$$

where w_t is the wage rate and "taxes" includes all kind of taxes that firms are in effect responsible for paying.

What are the asset-pricing implications of this type of model? First, the fundamental equation (1) is still valid. However, we now have a definition for capital income, consistent with a specific theory, that can be easily mapped into the U.S. national income and product account (NIPA) data.[2] The second prediction of the model is, as before, the familiar identity of market value and the value of all of the firm's assets (capital stock), after adjusting for dividend income taxes:

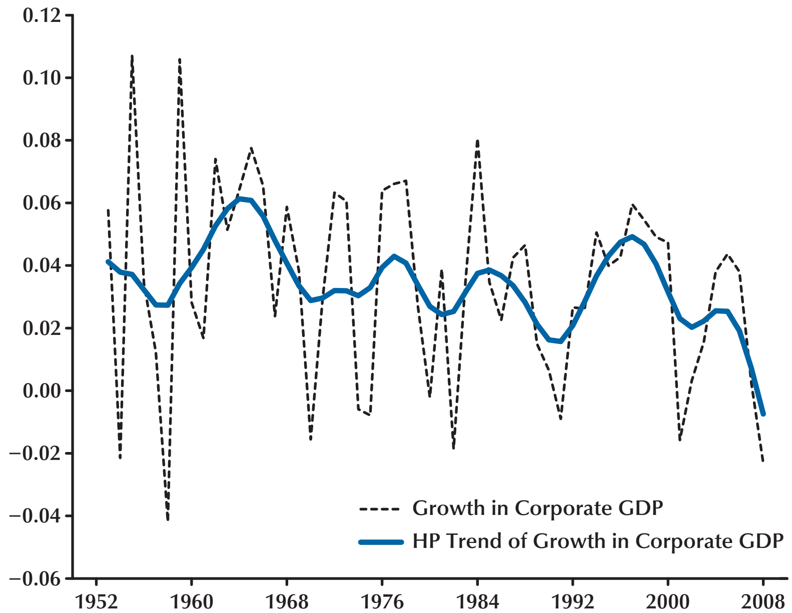
$$V_t = \left(1-\tau_t^d\right) d_t + \mathrm{E}_t\left[\left(1-\tau_{t+1}^V\right) p_{t+1} V_{t+1}\right].$$

This makes it clear that shareholders may obtain income from ownership of the firm in two different ways. The first is dividend payments, which is what the firm supposedly maximizes and that accrues to owners holding shares in perpetuity. The second way to obtain income is by selling equity shares, which may result in capital gains (or losses). Notice, finally, that standard measures of capital income equal shareholder income plus investment expenditures. This is consistent with our model. Investment expendi-

[2] Company-originated quarterly earnings are more comprehensive measures of capital income than dividends. However, we find that the results we report, when based on earnings, are very similar to those based on actual dividends.

Figure 7

Growth of Corporate GDP



tures are indeed a form of capital income since they may ultimately affect future dividends and the future value of the firm, which are both taken into account by the asset-pricing equation above. Put differently, shareholder total income in period t is the sum of the dividends received and of the (potential) capital gains accrued; the latter include (among other things) the value of period t's gross investment.

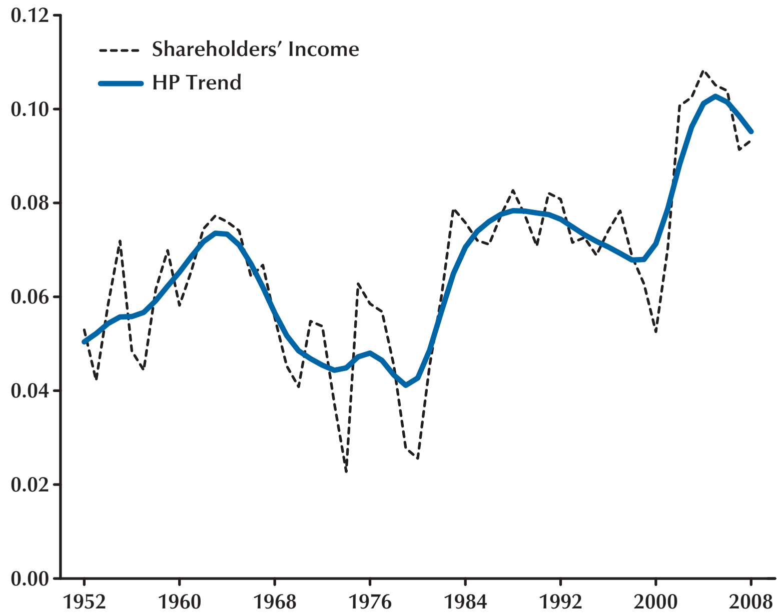
Because shareholder income is, by an accounting identity, equal to the fraction of corporate value added accruing to shareholders multiplied by corporate GDP, it is worth considering the two components separately to see whether their movements over time teach us anything useful. In standard macroeconomic models attention is focused, more often than not, on the time-series behavior of corporate GDP, while the share accruing to the owners of capital is taken as constant and paid little attention. This analytical choice is unfortunate since, as we will show later on, it may lead one to miss a fundamental factor affecting movements in stock market evaluations. Figures 7 and 8 report the two components separately. Notice that these are corporate GDP data and thus do not include the impact of changes in personal taxes in the net income accruing to shareholders.

There are various salient features in these data. First, corporate GDP growth fluctuates widely around an otherwise apparently stable long-run growth rate (with, possibly, a very modest downward trend in the latter period) and there are, really, only two decades of "major" (i.e., above-average) growth: the 1960s and the 1990s. Because these are also the two periods during which the market ratio rallied the most, one would be inclined to say that "roughly" the stock market captured the underlying long-run oscillations in payoff. The key word here, though, is "roughly"; in fact, very roughly as the subsequent quantitative analysis will show. Further, the recovery that ended almost two years ago was nothing spectacular: In terms of total corporate GDP growth, it was the *worst expansion* of the past 50 years!

Second, shareholder income as a fraction of total corporate value added (or capital income

Figure 8

Shareholders' Income (Relative to Corporate GDP)

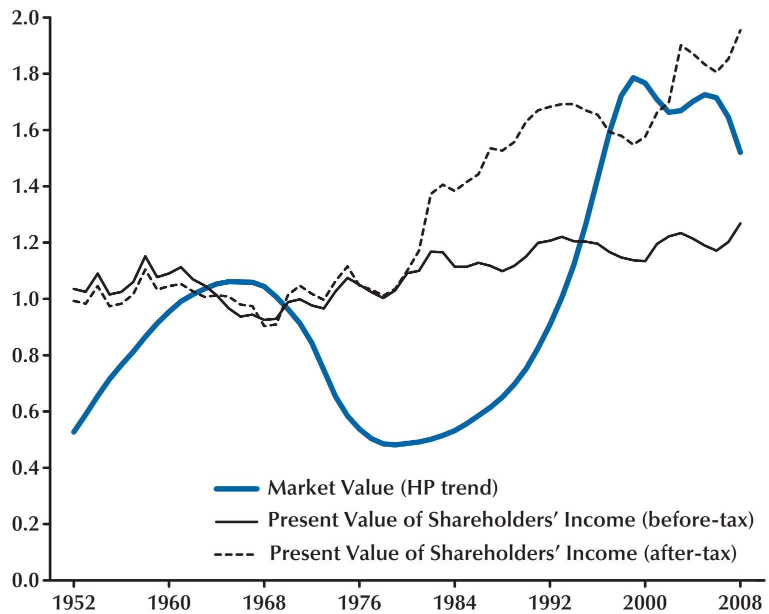


share) fluctuates widely over the sample period and has gone through the roof during the past decade. This fraction increased by 50 percent between 1953 and 1965, and then declined by 40 percent between 1966 and 1971 and remained at that level up to the early 1980s. For the period 1982-86, the data display a doubling of the capital income share, followed by a relative stabilization up to 2001, when the share increases sharply to unprecedented levels. By 2007 the capital income share is 40 percent higher than its 1983-2001 average and more than double what it was during the late 1960s and the 1970s. So much for the widespread assumption of long-run constancy of the capital share. Notice also that the long-run fluctuations we capture by means of the HP filter trend are dwarfed by the fluctuations taking place at business cycle frequencies: Factor shares in corporate income are anything but stable over time.

As a matter of fact, fluctuations in capital income are so large (in particular the 1982-2007 increase) that it becomes meaningless to perform a perfect foresight experiment symmetric to the ones noted above using the historically realized average growth rate of the capital share as the out-of-sample predictor for future dividends' growth. Because of the very large growth in the share of value added that is devoted to capital, the average growth rate of corporate capital income during the past 25 years is close to 6 percent, while corporate GDP grew on average by 3 percent. Assuming a permanent growth rate of 6 percent for corporate capital income implies that its share of corporate value added would become 100 percent a few decades in the future, which clearly makes no sense. A more reasonable experiment would then be to assume that, in the future, the capital share would remain constant at its average level during, say, the past 10 years, while corporate GDP grows into the infinite future at some reasonable rate. This is what we do, assuming that the capital share will remain at its average of the period 1998-2008 and, for consistency with our previous analysis, assuming that the current dividend tax will persist into the infinite future. The results are summarized in Figure 9. Interesting enough,

Figure 9

Present Value of Shareholders' Income versus Market Value (Relative to Corporate GDP)



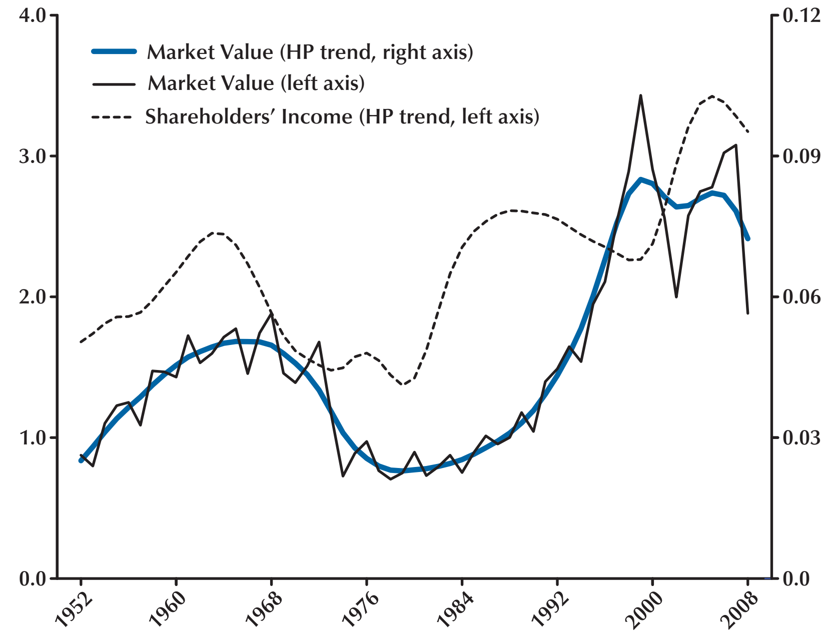
this reasonable modification does not make much of a difference, and the simulated market ratio resembles that of Figure 5.

It is also instructive to consider what the second equality implied by the theory, and by common sense, suggests: In normal circumstances the market value of a corporation should be equal to the replacement cost of its installed capital stock *plus* whatever organizational and intangible capital (patents, industrial secrets, and so on) the corporation controls. In principle, at least, a corporation should be able to sell its buildings, structures, and equipment (i.e., its tangible assets) on the market at roughly their replacement cost; hence, its market valuation should be lower than that of its tangible assets only in those special circumstances in which such tangible assets have been poorly invested and cannot be redirected to a different productive activity. Although this happens all the time at the level of the individual firm, one does not expect this to happen for roughly 32 of 56 years for the entire corporate sector. This is exactly what these data suggest happened!

Specifically, the time-series evolution of the K/Y ratio (at replacement value) as reported by NIPA, and summarized in Figure 2, moves a lot less than the market ratio; and it seems to be strongly negatively correlated with it as well. We observe very high investment levels (hence, of the capital stock in relation to output) in the mid-1970s, while equity values are persistently low. When the stock market trend inverts, so do investment and the K/Y ratio—hence, the low investment in the 1980s, with high equity values. Put differently, until about 1987, whenever an American firm purchased a piece of capital and installed it, that piece of capital immediately lost some value according to the stock market's prices. The common sense interpretation of this fact is that, for more than 30 years, the stock market considered the investment decisions of U.S. corporations to be "value reducing"! We call this a "puzzle" and, unless one is willing to theorize that "negative organizational capital" was accu-

Figure 10

Trends in Market Value and in Shareholders' Income (Relative to Corporate GDP)



mulated for three decades, this puzzle dwarfs the many others.

A recent literature offers an alternative interpretation of the previous facts. A high level of investment takes place when new and profitable technologies are first discovered, or adopted due to changes in the economic environment; profits come in later, when those technologies become fully operational and start producing fruitful results. Moreover, the fact that new technologies and new capital are introduced may render old capital obsolete, causing the market value of the latter to collapse (compare with Hobijn and Jovanovic, 2001, and Peralta-Alva, 2007). In this sense, one is tempted to read the high profits after 1982 as the return on the high level of investment of the 1970s. While this interpretation is perfectly reasonable and makes historical sense, it does require us to throw away a major tenet of most standard models: that is, that the stock market's prices embed an unbiased forecast of future corporate performances. The depressed stock market values of the mid-1970s and early 1980s did not seem to incorporate the payoffs of the ultimately successful investments of this period. Hence, these payoffs cannot be conceived as "expected." They happened, but the shareholders financing the high investment levels of those years were apparently unable to foresee the future gains those investments would have brought to them. This is puzzling.

During those years, instead, the share of capital in corporate income was at historical lows (Figure 10) and the market ratio seemed to, myopically, reflect more current miseries than future successes. The fact that it is hard to reconcile these observations with the theory is also emphasized by the technological change–driven explanations for the trends in the market ratio quoted above, because in those models the market ratio tends to recover much earlier than in the data. Notice that it is only after the mid-1980s, when firms' successes had been observed for some time and the capital share of corporate income has started to rise steadily, that the market ratio also picks up and starts reflecting current successes—or, maybe, forecasting future ones. A similar point can be made for pretty much every single major swing of the data we are considering: *Oscillations in the market ratio are anticipated by oscillations in the share of capital income in corporate GDP*. This observation suggests the need to look more carefully into the way in which shareholders forecast future performances and, in particular, into the role that current movements in the share of capital income play in determining shareholders' optimism or pessimism for the future.

BUILDING ON THE LESSONS FROM STANDARD MODELS

The main conclusions we derive from our previous analysis are as follows. First, actual dividends paid are too smooth to account for the key low-frequency trends in the market ratio; this remains true also when unexpected changes in the fiscal regime are taken into consideration. However, actual dividends paid are not necessarily what is priced by the stock market: While dividends paid are stable over time, we documented that the fraction of corporate value added captured by shareholders displays important fluctuations. Despite this adjustment, the asset valuation equation implied by different models under the perfect foresight hypothesis and a constant interest rate predicts a market ratio that is still too smooth relative to the data. Furthermore, we find that large classes of asset-pricing models where dividends are endogenously determined have some predictions that are hard to reconcile with the data. In particular, these models predict that market value should equal the value of the assets of the firm (after adjusting for taxes); in the data, these two series are negatively correlated—and most of the time the market value of the firm is lower than the value of the physical assets the firm controls! Finally, eyeball analysis suggests that movements in the share of capital in corporate income may be rough but consistent predictors of movements in the market ratio, an empirical regularity we now try to exploit.

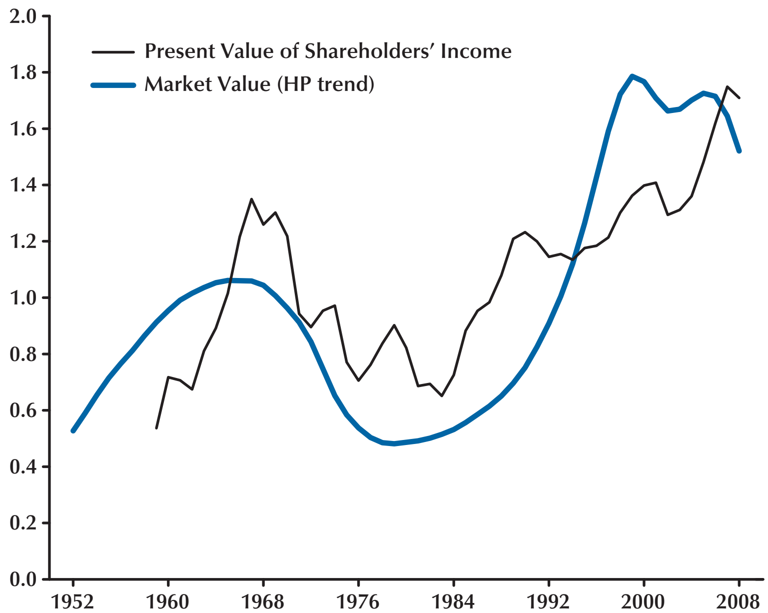
A delicate issue with all of our previous computations is the following: Pretending that in 1950 or in 1960, or even 1992 for that matter, shareholders could exactly forecast dividend payments in 2007 is clearly absurd. More important, the perfect foresight assumption typically results in a smooth series of predicted market values since all future fluctuations are foreseen and capitalized from the very beginning. Hence, a more reasonable "expectations formation" hypothesis needs to be introduced. While doing this opens a bottomless can of worms, this is an issue one must face squarely, especially if the study of past stock market behavior is supposed to shed some light on its current performance: What on earth drives shareholders' expectations? We consider this issue next.

In the past, people talked of "extrapolative expectations," arguing that—when forecasting the future in the absence of an understanding of the structural model driving the system—we look at trends in past data and extrapolate those trends over the relevant horizon. This happens, though, only when we have become convinced that they are, indeed, permanent trends and not just small and irrelevant blips. When evidence suggests that the trend has changed or reverted to old patterns, we accept it only after a while; but, once accepted, we tend to extrapolate it into the indefinite future. The problem, obviously, is how long is the "while" and how reasonable is it to assume that people extrapolate trends that cannot be sustained forever, such as the one we just noted to exist in the capital share of corporate value added during the past two decades? This is a hard question for which we do not have a good answer and that the learning literature has really never addressed. We will try, nevertheless, to make some practical progress along these lines.

We start by assuming that people extrapolate past trends forward, altering them as soon as "enough" evidence is obtained that the previous trend is no longer likely to persist. Following this idea, suppose agents use all the information available up to T periods in the past. We generate separate "forecasts" for the growth rate of corporate GDP and for the capital income share, using a weighted average on the observations for the

Figure 11

Present Value of Shareholders' Income (Distributed Lags) versus Market Value (Relative to Corporate GDP)



last $N < T$ periods. We focus, as before, on the classical trading strategy where infinitely long series of shareholder income are generated and used to predict market value. We compute "forecasts" based on the distributed lag equation,

$$X_{t+1} = \frac{1-\lambda}{1-\lambda^{N+1}} \sum_{i=0}^{N} \lambda^i X_{t-i},$$

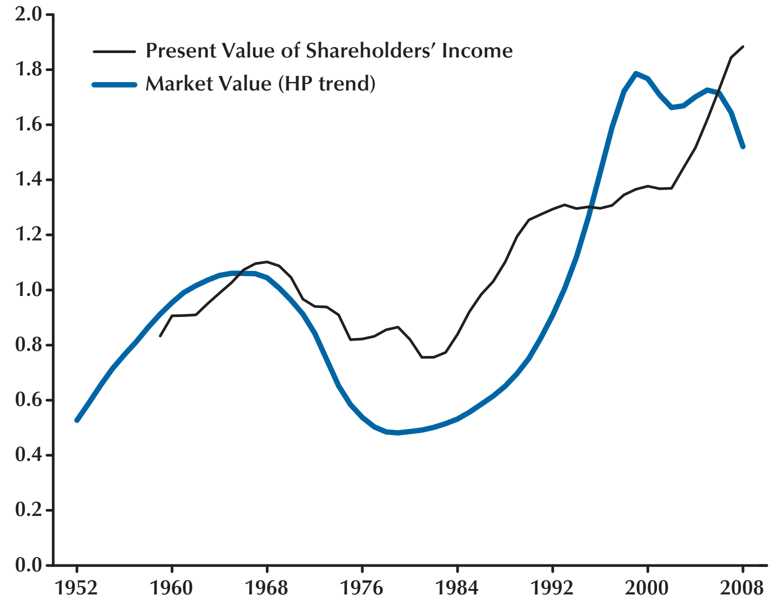
where X is the growth rate of each variable under consideration (in this case, the capital income share, corporate GDP growth rate, and dividend income tax rates).

We assume, as in our previous quantitative experiments, a constant discount factor of 7 percent and use the maximum number of lags possible at each moment in time (given our dataset). We then estimate the weights (one lambda for each time series being forecasted) in the distributed lag equation such that the sum of squared deviations between the theoretical market ratio and the data is minimized. (Further details can be found in our computational appendix.) The model's predictions and the data are summarized in Figure 11. As we can see, this simple approach can deliver a substantial improvement over the perfect foresight framework considered before. In particular, the predicted magnitudes for the 1960-68 increase, the mid-1970s' decline and stagnation that followed, and the ultimate recovery of market valuations are all comparable to those in the data. Notice, however, that the timing of the predicted changes in the trend of the market ratio tend to be off by a few years and that we cannot account for the large drop in market value of recent years either. Nevertheless, given the simplicity of our approach, we consider the predictions obtained by using this ad hoc form of extrapolative expectations interesting and worthy of more systematic pursuit.

Barsky and De Long (1993) follow a similar approach and conclude that dividend movements roughly account for the secular fluctuations in the U.S. stock market from the 1800s to 1993. These authors, however, abstract from changes in taxes

Figure 12

Present Value of Shareholders' Income (Distributed Lags and Constant Long-Run Growth) versus Market Value (Relative to Corporate GDP)

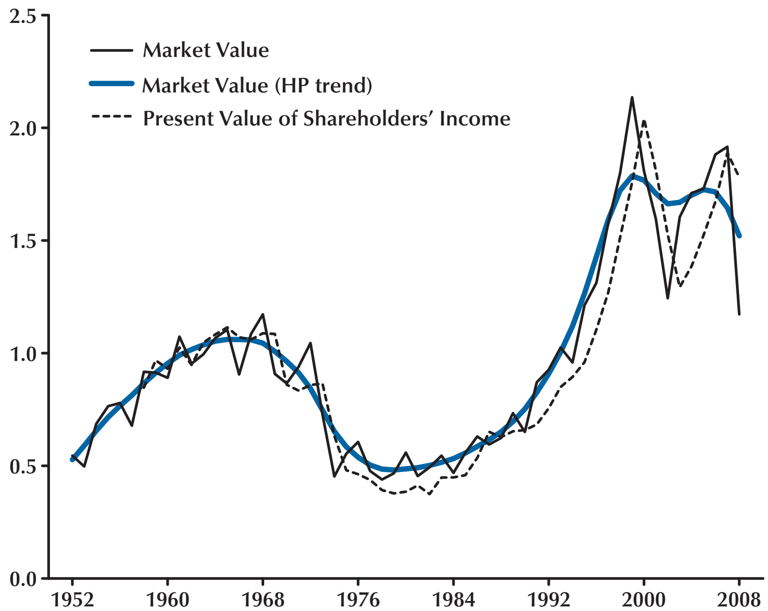


and use actual dividends paid by stock market firms in their analysis. As we illustrate above, however, actual dividends paid cannot account for the high stock market values observed after 1993. More important, these authors use a distributed lag equation similar to ours and estimate, period by period, a *permanent* growth rate of dividends. Such an estimation process implies (when applied to the actual data) that agents must expect dividends to grow at a rate permanently higher, or lower, than corporate GDP, which cannot really happen. Barsky and De Long's paper uses data up to the very early 1990s and hence does not have to face this puzzling prediction of their methodology, which is instead an implication of the past two decades of data. Furthermore, our quantitative analysis above has documented that, once one constrains the long-run growth rate of dividends to equal that of corporate GDP, it becomes impossible to account for observed stock market fluctuations. Our results in Figure 11 are computed based on forecasts of the HP trend of the growth rate of corporate GDP and, although this growth rate is far less volatile than dividends, may thus be subject to a similar criticism. To determine whether our results are driven by potentially unrealistic, permanent, forecasted values for the growth rate of corporate GDP, we consider a new experiment where a constant 3 percent growth rate for corporate GDP is assumed throughout. The results are summarized in Figure 12.

The fit of the model is still surprisingly good, but only up to a year ago: The drop in market ratio of the past year was unpredictable on the basis of the dividends performance observable up to 2007. It remains an open question as to what this methodology would predict a couple of years from now, when the substantial drop in capital income that took place in 2008 and 2009 will be reported in the data. This scenario notwithstanding, though, Figure 12 suggests that fluctuations in model-consistent shareholder income, and in taxes, can account for a large part of the secular movements of the U.S. stock market during the past 56 years *if* one is willing to assume that market participants use something akin to the distributed-lags fore-

Figure 13

Present Value of Shareholders' Income (Selling Shares After a Period) versus Market Value (Relative to Corporate GDP)



casting equation above in forming their expectations about the future. The key challenge for this simple framework seems to be accounting for the timing of the recovery during the mid-1980s and early 1990s. Capital income increased dramatically in the early 1980s. Dividend taxes declined substantially during the mid-1980s as well. According to the theory, these changes should have translated in a strong stock market recovery at the time. In the data, the recovery did start in the mid-1980s, but most of it did not take place until the mid-1990s.

Up to now, we have evaluated the asset-pricing equation of the basic model under a trading strategy of buy and hold (forever). Notice, however, that this fundamental asset-pricing equation holds for buy and hold, but it also implies that the value of the firm must equal the value of holding shares for *any* number of periods, T, and then selling and capturing the corresponding capital gains (or losses). Unfortunately, our current framework of analysis is not suited for studying the implications of the theory for short-term trading strategies. To understand why, observe that such analysis would require forecasting model–consistent capital income (as before), as well as future market values. But model-consistent capital income is a relatively small fraction of corporate GDP (between 6 and 10 percent), while market value is almost 10 times larger, between 50 and 160 percent of corporate GDP. Hence, for relatively short holding periods, fluctuations in the value of corporations predicted by the theory will be mostly driven by fluctuations in predicted market values. Indeed, when we apply the previous methods to derive the predictions of the asset-pricing equation for holding periods between three and five years, we obtain a very good fit not only for the HP trend of the market ratio, but also for the actual market ratio (Figure 13). The fact that the model matches well the HP trend of the market ratio, however, follows immediately from the fact that the HP trend of the market value (which is the focus of our analysis) is very persistent and predictable. To compute the predictions of the theory for period t's market value, we assume agents use all information available up to when the forecast is made, compute a new HP trend for market value, and use this market value trend to forecast (using the distributed lag equation above) future market values (and thus capital gains). Since an HP trend series that is updated continuously provides a good approximation (with a lag) to the underlying time series, our estimation method approximates well the actual market ratio (with a lag).

CONCLUDING REMARKS

We study fluctuations in the long-run trend of the ratio between stock market value and GDP for the U.S. corporate sector. According to economic theory, the market value of U.S. corporations should equal the *expected* present discounted value of the *future* flow of income and capital gains generated by this sector. This prediction of the theory is frequently tested assuming *perfect foresight* on *actual* dividends paid. Actual dividends are very smooth and their movements cannot account for stock market trends, even in the long run. Many researchers consider this a puzzle. We find that a measure of *model-consistent* dividends fluctuates much more than actual dividends paid. More important, fluctuations in model-consistent dividends are positively correlated with stock market fluctuations. We illustrate that the perfect foresight assumption, by construction, predicts a very smooth present value of model-consistent dividends, and thus a very smooth market ratio, even when dividends fluctuate a lot. Theory does not require and does not imply that individuals and firms have perfect foresight, however; it simply requires and predicts that individuals will use all available information optimally (that is, as well as they can) to form their expectations of future movements in capital income. We then evaluate the theory under the assumption that all available (but no future) information is used in an extrapolative expectations format to forecast future dividend payments. We use a distributed lag equation to do so. We find that the present value of dividends, computed in this way, is much more consistent with the data. Apart from the obvious question of what, other than wisdom after the fact, may justify or explain the particular choice of forecasting rule made by market participants, our analysis leaves open an important puzzle: The value of corporations should equal the value of their tangible and intangible assets, while in the data the two series seem to be negatively correlated and remain persistently apart from each other.

REFERENCES

Barsky, Robert B. and De Long, James B. "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics*, May 1993, *108*(2), pp. 291-311.

Bian, Jiangze. *Three Essays on Macroeconomics and Asset Prices*. Ph.D. Dissertation, University of Miami, 2007.

Boldrin, Michele; Christiano, Lawrence J. and Fisher, Jonas D.M. "Habit Persistence, Asset Returns, and the Business Cycle." *American Economic Review*, March 2001, *91*(1), pp. 149-66.

Campbell, John Y. "Consumption-Based Asset Pricing," in G.M. Constantinides, M. Harris and R.M. Stulz, eds., *Handbook of the Economics of Finance*. New York: Elsevier, 2003, pp. 803-997.

Easterbrook, Frank H. "Two Agency-Cost Explanation of Dividends." *American Economic Review*, September 1984, *74*(4), pp. 650-59.

Feldstein, Martin and Green, Jerry. "Why Do Companies Pay Dividends?" *American Economic Review*, 1983, *73*(1), pp. 17-30.

Hobijn, Bart and Jovanovic, Bojan. "The Information-Technology Revolution and the Stock Market: Evidence." *American Economic Review*, December 2001, *91*(5), pp. 1203-20.

McGrattan, Ellen R. and Prescott, Edward C. "Taxes, Regulation, and the Value of U.S. and U.K. Corporations." *Review of Economic Studies*, July 2005, *72*(252), pp. 767-96.

Peralta-Alva, Adrian. "The Information Technology Revolution and the Puzzling Trends in Tobin's Average 'q.'" *International Economic Review*, August 2007, *48*(3), pp. 929-51.

Ravn, Morten O. and Uhlig, Harald. "On Adjusting the HP-Filter for the Frequency of Observations." CEPR Discussion Paper No. 2858, Centre for Economic Policy Research, 2001.

Shiller, Robert J. *Irrational Exuberance*. Princeton, NJ: Princeton University Press, 2005.

DATA APPENDIX

- The *market value of corporations* is based on quarterly data in "Table L.213: Issues at Market Value" from the Flows of Funds Account of the United States from the Federal Reserve. In a given year, we take the last quarter of the quarterly market value to create the annual level.
- *Corporate value added* (*or corporate GDP*) is "Table 1.14: Gross Value Added of Corporate Business" from the National Income and Product Accounts (NIPA) published by the Bureau of Economic Analysis (BEA). Data on corporate businesses is obtained NIPA.
- The *replacement value of corporate capital* is the sum of nonresidential and residential tangible corporate fixed assets, measured at current cost, as reported in the Standard Fixed Asset Tables (Tables 4.1 and 5.1) published by the BEA.
- *Dividend tax rates* up to 1998 are from the data appendix of McGrattan and Prescott (2005), and the rate from 1999-2005 from Bian (2007), who followed the methodology of McGrattan and Prescott. The rate for 2006-2008 was assumed equal to that for 2005.
- Both *compensation of employees* and *corporate income taxes* were taken from "Table 1.14: Gross Value Added of Domestic Corporate Business" in Domestic Product and Income from NIPA. Compensation of employees is line 4, while taxes are the sum of corporate income taxes (line 12) and taxes on production and import less subsidies (line 7). Corporate *gross investment* is from the Standard Fixed Asset Tables (Tables 4.7 and 5.7), and it is equal to the sum of Investment in Private Nonresidential and Residential Fixed Assets of U.S. corporations.

Transformations on the Data

Except for tax rates, we divide all the series by the GDP implicit price deflator (2000 = 100) before we begin our analysis. In the annual series to which we apply the HP trend, we use a smoothing parameter of 6.25, recommended by Ravn and Uhlig (2001). To estimate Figure 13, we use the HP trend of the quarterly market value. In this case, we use a smoothing parameter of 1600.

COMPUTATIONAL APPENDIX

This section describes the algorithm used to compute the market value forecasts summarized in Figures 11 and 12.

We start by constructing forecasts for corporate GDP growth, the share of dividends in corporate GDP, and dividend tax rates. These forecasts are based on a distributed lag equation. In particular, standing at period t, we compute the period $t+1$ forecasted value for variable X as follows:

$$\tilde{X}_{t+1|t} = \frac{1-\lambda_X}{1-\lambda_X^N} \sum_{i=0}^{N} \lambda_X^i X_{t-i|t}.$$

Here, λ_x denotes the weight of past observations, and N the number of lags included in the forecast. Let $(\tilde{g}_{t+k/t}, \tilde{\tau}_{d,t+k|t}, \tilde{s}_{d,t+k/t})$ denote the resulting period t–forecasted sequences for the growth rate of corporate GDP, the tax rate on dividend income taxes, and the dividend share of corporate GDP.

Then, our forecast at period t for dividends at t+k is given by:

$$\tilde{d}_{t+k|t} = \left(1-\tilde{\tau}_{d,t+k|t}\right)\tilde{s}_{d,t+k|t} GDPcorp_{HP-trend,t} \prod_k \left(1+\tilde{g}_{t+k|t}\right).$$

We compute the predictions of the model using *only* information available up to the period when the forecast is made. This entails computing a new set of HP trends from the data, as well as new out-of-sample forecasted sequences for dividends, GDP growth, and taxes for each given year.

The market values reported in Figure 11 are thus given by the present value of dividends condition:

$$V_t = \sum_{k=1}^{\infty} \frac{\tilde{d}_{t+k|t}}{\left(1+r\right)^k}.$$

Finally, the values of the weights in the distributed lag equations above, $(\lambda_d, \lambda_{\tau_d})$, where chosen to minimize the square sum of residuals between forecasted and observed market values (as ratios to corporate GDP). The values we used are $(\lambda_d = 0.65, \lambda_{\tau_d} = 0.79)$.

The results reported in Figure 12 are obtained following a symmetric procedure, where we replace the last term of the dividend forecasting equation above,

$$\prod_k \left(1+\tilde{g}_{t+k|t}\right),$$

by $(1 + 0.03)^k$.

Federal Reserve Bank of St. Louis *Review*, Annual Index, 2009

Carlos Hamilton Araujo, Seppo Honkapohja, and James Bullard
Panel Discussion

Commentaries by:

Rodolfo E. Manuelli
Jon Faust
Gregor W. Smith
Robert J. Tetlow
Ellis W. Tallman
Xiaodong Zhu
Dean Croushore

SEPTEMBER/OCTOBER, PART 1

Raghuram G. Rajan
"The Credit Crisis and Cycle-Proof Regulation"

James Bullard, Christopher J. Neely, and David C. Wheelock
"Systemic Risk and the Financial Crisis: A Primer"

David C. Wheelock and Mark E. Wohar
"Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature"

Rodolphe Blavy and Luciana Juvenal
"Mexico's Integration into NAFTA Markets: A View from Sectoral Real Exchange Rates"

SEPTEMBER/OCTOBER, PART 2

Edward Nelson
"Milton Friedman and U.K. Economic Policy: 1938-1979"

Marlene Amstad and Andreas M. Fischer
"Do Macroeconomic Announcements Move Inflation Forecasts?"

Don H. Kim
"Challenges in Macro-Finance Modeling"

Rocco Ciciretti, Gerald P. Dwyer, and Iftekhar Hasan
"Investment Analysts' Forecasts of Earnings"

NOVEMBER/DECEMBER

Jeffrey P. Cohen, Cletus C. Coughlin, and Lesli S. Ott
"Auctions as a Vehicle to Reduce Airport Delays and Achieve Value Capture"

Richard G. Anderson and Charles S. Gascon
"The Commercial Paper Market, the Fed, and the 2007-2009 Financial Crisis"

Dan A. Black, Natalia A. Kolesnikova, and Lowell J. Taylor
"Local Price Variation and Labor Supply Behavior"

Michele Boldrin and Adrian Peralta-Alva
"What Happened to the U.S. Stock Market? Accounting for the Past 50 Years"

Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166-0442